CERTIFICATE OF MAILING BY "EXPRESS MAIL"

EXPRESS MAIL LABEL NUMBER: EV 318 738 041 US

DATE OF DEPOSIT: July 14, 2003

I HEREBY CERTIFY THAT THIS PAPER OR FEE IS BEING DEPOSITED WITH THE UNITED STATES POSTAL SERVICE "EXPRESS MAIL POST OFFICE TO ADDRESSEE" SERVICE UNDER 37 CFR 1.10 ON THE DATE INDICATED ABOVE AND IS ADDRESSED TO MAIL STOP PATENT APPLICATION, COMMISSIONER FOR PATENTS, P.O. BOX 1450, ALEXANDRIA, VA 22313-1450.

MIKHAIL BAYLEY

DIVISIONAL

APPLICATION

for

UNITED STATES LETTERS PATENT

on

LONG WAVELENGTH ENGINEERED FLUORESCENT PROTEINS

by

Rebekka M. Wachter

and

S. James Remington

Sheets of Drawings: Sixty Two (62)
Docket No.: REGEN1250-7

Lisa A. Haile, J.D., Ph.D.
Gray Cary Ware & Freidenrich LLP
4365 Executive Drive, Suite 1100
San Diego, California 92121-2133

PATENT

Attorney Docket No. REGEN1250-5

LONG WAVELENGTH ENGINEERED FLUORESCENT PROTEINS

BACKGROUND OF THE INVENTION

This a continuation in part of Application No 08/974,737, filed 11/19/97, now allowed, which is a continuation of Application No 08/911,825, filed 8/15/97, now issued as US patent No. 6,054,321, which is a continuation in part of Application No. 08/706,408, filed 8/30/96, now allowed, which claims the benefit of the earlier filing date of a United States provisional patent application serial number 60/024,050 filed on August 16, 1996 each of which are herein incorporated by reference.

This invention was made in part with Government support under grant no. MCB 9418479 awarded by the National Science Foundation. The Government may have rights in this invention.

Fluorescent molecules are attractive as reporter molecules in many assay systems because of their high sensitivity and ease of quantification. Recently, fluorescent proteins have been the focus of much attention because they can be produced *in vivo* by biological systems, and can be used to trace intracellular events without the need to be introduced into the cell through microinjection or permeablization. The green fluorescent protein of *Aequorea victoria* is particularly interesting as a fluorescent protein. A cDNA for the protein has been cloned. (D.C. Prasher et al., "Primary structure of the *Aequorea victoria* green-fluorescent protein," *Gene* (1992) 111:229-33.) Not only can the primary amino acid sequence of the protein be expressed from the cDNA, but the expressed protein can fluoresce. This indicates that the protein can undergo the cyclization and oxidation believed to be necessary for fluorescence. *Aequorea* green fluorescent protein ("GFP") is a stable, proteolysis-resistant single chain of 238 residues and has two absorption maxima at around 395 and 475 nm. The relative amplitudes of these two peaks is sensitive to environmental factors (W. W. Ward. *Bioluminescence and Chemiluminescence* (M.

A. DeLuca and W. D. McElroy, eds) Academic Press pp. 235-242 (1981); W. W. Ward & S. H. Bokman *Biochemistry* 21:4535-4540 (1982); W. W. Ward et al. *Photochem. Photobiol.* 35:803-808 (1982)) and illumination history (A. B. Cubitt et al. *Trends Biochem. Sci.* 20:448-455 (1995)), presumably reflecting two or more ground states. Excitation at the primary absorption peak of 395 nm yields an emission maximum at 508 nm with a quantum yield of 0.72-0.85 (O. Shimomura and F.H. Johnson *J. Cell. Comp. Physiol.* 59:223 (1962); J. G. Morin and J. W. Hastings, *J. Cell. Physiol.* 77:313 (1971); H. Morise et al. *Biochemistry* 13:2656 (1974); W. W. Ward *Photochem. Photobiol. Reviews* (Smith, K. C. ed.) 4:1 (1979); A. B. Cubitt et al. *Trends Biochem. Sci.* 20:448-455 (1995); D. C. Prasher *Trends Genet.* 11:320-323 (1995); M. Chalfie *Photochem. Photobiol.* 62:651-656 (1995); W. W. Ward. *Bioluminescence and Chemiluminescence* (M. A. DeLuca and W. D. McElroy, eds) Academic Press pp. 235-242 (1981); W. W. Ward & S. H. Bokman *Biochemistry* 21:4535-4540 (1982); W. W. Ward et al. *Photochem. Photobiol.* 35:803-808 (1982)). The fluorophore results from the autocatalytic cyclization of the polypeptide backbone between residues Ser^{65} and Gly^{67} and oxidation of the -ß bond of Tyr^{66} (A. B. Cubitt et al. *Trends Biochem. Sci.* 20:448-455 (1995); C. W. Cody et al. *Biochemistry* 32:1212-1218 (1993); R. Heim et al. *Proc. Natl. Acad. Sci.* USA 91:12501-12504 (1994)). Mutation of Ser^{65} to Thr (S65T) simplifies the excitation spectrum to a single peak at 488 nm of enhanced amplitude (R. Heim et al. *Nature* 373:664-665 (1995)), which no longer gives signs of conformational isomers (A. B. Cubitt et al. *Trends Biochem. Sci.* 20:448-455 (1995)).

Fluorescent proteins have been used as markers of gene expression, tracers of cell lineage and as fusion tags to monitor protein localization within living cells. (M. Chalfie et al., "Green fluorescent protein as a marker for gene expression," *Science* 263:802-805; A.B. Cubitt et al., "Understanding, improving and using green fluorescent proteins," *TIBS 20*, November 1995, pp. 448-455. U.S. patent 5,491,084; M. Chalfie and D. Prasher. Furthermore, engineered versions of *Aequorea* green fluorescent protein have been identified that exhibit altered fluorescence characteristics, including altered excitation and emission maxima, as well as excitation and emission spectra of different shapes. (R. Heim et al., "Wavelength

mutations and posttranslational autoxidation of green fluorescent protein," *Proc. Natl. Acad. Sci. USA*, (1994) 91:12501-04; R. Heim et al., "Improved green fluorescence," *Nature* (1995) 373:663-665.)

A second class of applications rely on GFP as a specific indicator of some cellular property, and hence depend on the particular spectral characteristics of the variant employed. For recent reviews on GFP variants and their applications, see (Palm & Wlodawer, 1999; Tsien, 1998) , and for a review volume on specialized applications, see (Sullivan & Kay, 1999) . Biosensor applications include the use of differently colored GFPs for fluorescence resonance energy transfer (FRET) to monitor protein-protein interactions (Heim, 1999) or Ca2+ concentrations (Miyawaki *et al.*, 1999) , and receptor insertions within GFP surface loops to monitor ligand binding (Baird *et al.*, 1999; Doi & Yanagawa, 1999) .

The fluorescence emission of a number of variants is highly sensitive to the acidity of the environment (Elsliger *et al.*, 1999; Wachter *et al.*, 1998) . Hence, one particularly successful application of green fluorescent protein (GFP) as a visual reporter in live cells has been the determination of organelle or cytosol pH (Kneen *et al.*, 1998; Llopis *et al.*, 1998; Miesenbock *et al.*, 1998; Robey *et al.*, 1998) . The two chromophore charge states have been found to be relevant to the pH sensitivity of the intact protein, and have been characterized crystallographically in terms of conformational changes in the vicinity of the phenolic end (Elsliger *et al.*, 1999) , and spectroscopically using Raman studies (Bell *et al.*, 2000) . The neutral form of the chromophore, band A, absorbs around 400 nm in most variants, whereas the chromophore anion with the phenolic end deprotonated (band B) absorbs in the blue to green, depending on the particular mutations in the vicinity of the chromophore. WT GFP exhibits spectral characteristics that are consistent with two ground states characterized by a combination of bands A and B, the ratio of which is relatively invariant between pH 6 and 10 (Palm & Wlodawer, 1999; Ward *et al.*, 1982) . It has been suggested that an internal equilibrium exists where a proton is shared between the chromophore phenolate and the carboxylate of Glu222 over a broad range of pH (Brejc *et al.*, 1997; Palm *et al.*, 1997) . Recent electrostatic calculations support this model (Scharnagl *et al.*, 1999) , and estimate the theoretical pK_a for complete

chromophore deprotonation to be about 13, consistent with the observation of a doubling of emission intensity at pH 11-12 (Bokman & Ward, 1981; Palm & Wlodawer, 1999) .

In contrast to WT GFP, the chromophore of most variants titrates with a single pK_a. The color emission and the chromophore pK_a are strongly modulated by the protein surroundings (Llopis *et al.*, 1998) . Glu222 is completely conserved among GFP homologs (Matz *et al.*, 1999) , and its substitution by a glutamine has been shown to dramatically reduce efficiency of chromophore generation (Elsliger *et al.*, 1999) . Protonation of Glu222 in S65T and in GFPs containing the T203Y mutation (YFPs) is generally thought to be responsible for lowering the chromophore pKa from that of WT to about 5.9 in GFP S65T (Elsliger *et al.*, 1999; Kneen *et al.*, 1998) , and 5.2 -5.4 in YFP (GFP S65G/V68L/S72A/T203Y) (Ormo *et al.*, 1996; Wachter & Remington, 1999). In the YFPs, it is thought that the crystallographically identified stacking interaction of the chromophore with Tyr203 is largely responsible for the spectral red-shift (Wachter *et al.*, 1998) .

Unlike other variants, we have discovered that the YFP chromophore pK_a shows a strong dependence on the concentration of certain small anions such as chloride (Wachter & Remington, 1999) , and increases in pK_a from about 5.2 to 7.0 in the presence of 140 mM NaCl (Elsliger *et al.*, 1999) . This sensitivity can be exploited to enable the creation of novel GFPs as biosensors to measure ions present both in the cytoplasm or in cellular compartments (Wachter & Remington, 1999) within living cells. The present invention includes the creation and use of novel GFP variants that permit the fluorescent measurement of a variety of ions, including halides such as chloride and iodide. These properties add variety and utility to the arsenal of biologically based fluorescent indicators. There is a need for engineered fluorescent proteins with varied fluorescent properties and with the ability to respond to ion concentrations via a change in fluorescence characteristics.

BRIEF DESCRIPTION OF THE DRAWINGS

Figs. 1A-1B. (A) Schematic drawing of the backbone of GFP produced by Molscript (J.P. Kraulis, *J. Appl. Cryst.*, 24:946 (1991)). The chromophore is shown as a ball and stick model. (B) Schematic drawing of the overall fold of GFP. Approximate residue numbers mark the beginning and ending of the secondary structure elements.

Figs. 2A-2C. (A) Stereo drawing of the chromophore and residues in the immediate vicinity. Carbon atoms are drawn as open circles, oxygen is filled and nitrogen is shaded. Solvent molecules are shown as isolated filled circles. (B) Portion of the final $2F_o$-F_c electron density map contoured at 1.0 σ, showing the electron density surrounding the chromophore. (C) Schematic diagram showing the first and second spheres of coordination of the chromophore. Hydrogen bonds are shown as dashed lines and have the indicated lengths in Å. Inset: proposed structure of the carbinolamine intermediate that is presumably formed during generation of the chromophore.

Fig. 3 depicts the nucleotide sequence (SEQ ID NO:1) and deduced amino acid sequence (SEQ ID NO:2) of an *Aequorea* green fluorescent protein.

Figs. 4A-B depict the nucleotide sequence (SEQ ID NO:3) and deduced amino acid sequence (SEQ ID NO:4) of the engineered *Aequorea*-related fluorescent protein S65G/S72A/T203Y utilizing preferred mammalian codons and optimal Kozak sequence.

Figs. 5A to 5AT present the coordinates for the crystal structure of *Aequorea*-related green fluorescent protein S65T.

Fig. 6 shows the fluorescence excitation and emission spectra for engineered fluorescent proteins 20A and 10C (Table F). The vertical line at 528 nm compares the emission maxima of 10C, to the left of the line, and 20A, to the right of the line.

Fig 7. Shows absorbance scans of YFP at varying NaCl concentration and constant pH 6.4, buffered with 20 mM MES (–O– 0 mM NaCl, –∇– 15 mM NaCl, –□– 50mM NaCl, –◇– 100 mM NaCl, and –Δ– 400 mM NaCl). Band A

corresponds to the neutral form of the chromophore (λ_{max}=392nm), and band B corresponds to the chromophore anion (λ_{max}=514nm).

Fig 8. Shows normalized fluorescence emission of (a) YFP and (b) YFP-H148Q, as a function of pH and [Cl^-] at constant ionic strength of 150 mM. The pH was controlled with 20 mM TAPS pH 8.0 (O), 20 mM HEPES pH 7.5 (Δ), 20 mM PIPES pH 7.0 (◇), and 20 mM MES pH 6.5 (▽) and pH 6.0 ( ). Figure 8(b) also includes the fluorescence emission of YFP-H148Q as a function of [I^-] at pH 7.5 (*). Potassium chloride (or iodide) was added to the indicated concentration, and the ionic strength was adjusted to 150 mM with potassium gluconate. The samples, containing approximately 0.01 mg/ml protein, were excited at 514 nm, and emission intensity was determined at 528 nm.

Fig. 9 Shows a stereoview of the $2F_o$-F_c electron-density map of the YFP-H148Q chromophore, Tyr203, Arg96, Gln69, and the buried iodide after refinement. The 2.1 Å resolution map was contoured at +1 standard deviation. This figure was drawn by the program BOBSCRIPT.

Fig. 10 Shows a schematic diagram showing all residues that contain atoms within 5 Å of the buried iodide in the crystal structure of YFP-H148Q (iodide soak).

Fig. 11 Shows a stereoview of an overlay of a subset of residues lining the anion binding cavity of YFP-H148Q, with and without iodide (iodide-bound structure, grey; apo-structure, black). The iodide is represented by the center sphere. This figure was drawn by the program MOLSCRIPT (Kraulis, 1991).

Fig 12. Shows a schematic diagram of the immediate chromophore environment of YFP-H148Q in the (a) apo-structure, and (b) iodide-bound structure.

Fig. 13 Shows a stereoview of the solvent-accessible surface of the iodide-bound YFP-H148Q structure, calculated using a 1.4 Å probe radius. The surface was calculated after deleting all water molecules and the iodide. The chromophore and all surface segments in contact with the chromophore are also shown. The outer surface

of the protein is along the left edge of the figure. This figure was generated using the program MidasPlus™ (UCSF, 1994).

Fig. 14 Shows the backbone atom trace of β-strands 7 and 8 of YFP, the apo-structure of YFP-H148Q, and the iodide-bound structure of YFP-H148Q. The side chain of His148 (YFP) and Gln148 (YFP-H148Q), and a few water molecules are also shown. The dashed lines represent possible hydrogen bonds.

Fig. 15 Shows YFP chromophore pK_a as a function of halide concentration (–◇– fluoride, –∇– iodide, –Δ– chloride, –O– bromide). The chromophore pK_a was estimated from absorbance scans at varying halide concentrations (see Materials and Methods). The data were curve-fit to equation 1 (see text).

SUMMARY OF THE INVENTION

This invention provides functional engineered fluorescent proteins with varied fluorescence characteristics that can be easily distinguished from currently existing green and blue fluorescent proteins. Such engineered fluorescent proteins enable the simultaneous measurement of two or more processes within cells and can be used as fluorescence energy donors or acceptors, as well as biosensors for detecting anions. Longer wavelength engineered fluorescent proteins are particularly useful because photodynamic toxicity and auto-fluorescence of cells are significantly reduced at longer wavelengths. In particular, the introduction of the substitution T203X, wherein X is an aromatic amino acid, results in an increase in the excitation and emission wavelength maxima of *Aequorea*-related fluorescent proteins.

In one aspect, this invention provides a nucleic acid molecule comprising a nucleotide sequence encoding a functional engineered fluorescent protein whose amino acid sequence is substantially identical to the amino acid sequence of *Aequorea* green fluorescent protein (SEQ ID NO:2) and which differs from SEQ ID NO:2 by at least an amino acid substitution located no more than about 0.5 nm from the chromophore of the engineered fluorescent protein, wherein the substitution alters the electronic environment of the chromophore, whereby the functional engineered

fluorescent protein has a different fluorescent property than *Aequorea* green fluorescent protein.

In one aspect this invention provides a nucleic acid molecule comprising a nucleotide sequence encoding a functional engineered fluorescent protein whose amino acid sequence is substantially identical to the amino acid sequence of *Aequorea* green fluorescent protein (SEQ ID NO:2) and which differs from SEQ ID NO:2 by at least a substitution at T203 and, in particular, T203X, wherein X is an aromatic amino acid selected from H, Y, W or F, said functional engineered fluorescent protein having a different fluorescent property than *Aequorea* green fluorescent protein. In one embodiment, the amino acid sequence further comprises a substitution at S65, wherein the substitution is selected from S65G, S65T, S65A, S65L, S65C, S65V and S65I. In another embodiment, the amino acid sequence differs by no more than the substitutions S65T/T203H; S65T/T203Y; S72A/F64L/S65G/T203Y; S65G/V68L/Q69K/S72A/T203Y; S72A/S65G/V68L/T203Y; S65G/S72A/T203Y; or S65G/S72A/T203W. In another embodiment, the amino acid sequence further comprises a substitution at Y66, wherein the substitution is selected from Y66H, Y66F, and Y66W. In another embodiment, the amino acid sequence further comprises a mutation from Table A. In another embodiment, the amino acid sequence further comprises a folding mutation. In another embodiment, the nucleotide sequence encoding the protein differs from the nucleotide sequence of SEQ ID NO:1 by the substitution of at least one codon by a preferred mammalian codon. In another embodiment, the nucleic acid molecule encodes a fusion protein wherein the fusion protein comprises a polypeptide of interest and the functional engineered fluorescent protein.

In another aspect, this invention provides a nucleic acid molecule comprising a nucleotide sequence encoding a functional engineered fluorescent protein whose amino acid sequence is substantially identical to the amino acid sequence of *Aequorea* green fluorescent protein (SEQ ID NO:2) and which differs from SEQ ID NO:2 by at least an amino acid substitution at L42, V61, T62, V68, Q69, Q94, N121, Y145, H148, V150, F165, I167, Q183, N185, L220, E222 (not E222G), or V224, said

functional engineered fluorescent protein having a different fluorescent property than *Aequorea* green fluorescent protein. In one embodiment, amino acid substitution is:

L42X, wherein X is selected from C, F, H, W and Y,
V61X, wherein X is selected from F, Y, H and C,
T62X, wherein X is selected from A, V, F, S, D, N, Q, Y, H and C,
V68X, wherein X is selected from F, Y and H,
Q69X, wherein X is selected from K, R, E and G,
Q94X, wherein X is selected from D, E, H, K and N,
N121X, wherein X is selected from F, H, W and Y,
Y145X, wherein X is selected from W, C, F, L, E, H, K and Q,
H148X, wherein X is selected from F, Y, N, K, Q and R,
V150X, wherein X is selected from F, Y and H,
F165X, wherein X is selected from H, Q, W and Y,
I167X, wherein X is selected from F, Y and H,
Q183X, wherein X is selected from H, Y, E and K,
N185X, wherein X is selected from D, E, H, K and Q,
L220X, wherein X is selected from H, N, Q and T,
E222X, wherein X is selected from N and Q, or
V224X, wherein X is selected from H, N, Q, T, F, W and Y.

In a further aspect, this invention provides an expression vector comprising expression control sequences operatively linked to any of the aforementioned nucleic acid molecules. In a further aspect, this invention provides a recombinant host cell comprising the aforementioned expression vector.

In another aspect, this invention provides a functional engineered fluorescent protein whose amino acid sequence is substantially identical to the amino acid sequence of *Aequorea* green fluorescent protein (SEQ ID NO:2) and which differs from SEQ ID NO:2 by at least an amino acid substitution located no more than about 0.5 nm from the chromophore of the engineered fluorescent protein, wherein the substitution alters the electronic environment of the chromophore, whereby the

functional engineered fluorescent protein has a different fluorescent property than *Aequorea* green fluorescent protein.

In another aspect, this invention provides a functional engineered fluorescent protein whose amino acid sequence is substantially identical to the amino acid sequence of *Aequorea* green fluorescent protein (SEQ ID NO:2) and which differs from SEQ ID NO:2 by at least the amino acid substitution at T203, and in particular, T203X, wherein X is an aromatic amino acid selected from H, Y, W or F, said functional engineered fluorescent protein having a different fluorescent property than *Aequorea* green fluorescent protein. In one embodiment, the amino acid sequence further comprises a substitution at S65, wherein the substitution is selected from S65G, S65T, S65A, S65L, S65C, S65V and S65I. In another embodiment, the amino acid sequence differs by no more than the substitutions S65T/T203H; S65T/T203Y; S72A/F64L/S65G/T203Y; S72A/S65G/V68L/T203Y; S65G/V68L/Q69K/S72A/T203Y; S65G/S72A/T203Y; or S65G/S72A/T203W. In another embodiment, the amino acid sequence further comprises a substitution at Y66, wherein the substitution is selected from Y66H, Y66F, and Y66W. In another embodiment, the amino acid sequence further comprises a folding mutation. In another embodiment, the engineered fluorescent protein is part of a fusion protein wherein the fusion protein comprises a polypeptide of interest and the functional engineered fluorescent protein.

In another aspect this invention provides a functional engineered fluorescent protein whose amino acid sequence is substantially identical to the amino acid sequence of *Aequorea* green fluorescent protein (SEQ ID NO:2) and which differs from SEQ ID NO:2 by at least an amino acid substitution at L42, V61, T62, V68, Q69, Q94, N121, Y145, H148, V150, F165, I167, Q183, N185, L220, E222, or V224, said functional engineered fluorescent protein having a different fluorescent property than *Aequorea* green fluorescent protein.

In another aspect, this invention provides a fluorescently labelled antibody comprising an antibody coupled to any of the aforementioned functional engineered fluorescent proteins. In one embodiment, the fluorescently labelled antibody is a

fusion protein wherein the fusion protein comprises the antibody fused to the functional engineered fluorescent protein.

In another aspect, this invention provides a nucleic acid molecule comprising a nucleotide sequence encoding an antibody fused to a nucleotide sequence encoding a functional engineered fluorescent protein of this invention.

In another aspect, this invention provides a fluorescently labelled nucleic acid probe comprising a nucleic acid probe coupled to a functional engineered fluorescent protein whose amino acid sequence of this invention. The fusion can be through a linker peptide.

In another aspect, this invention provides a method for determining whether a mixture contains a target comprising contacting the mixture with a fluorescently labelled probe comprising a probe and a functional engineered fluorescent protein of this invention; and determining whether the target has bound to the probe. In one embodiment, the target molecule is captured on a solid matrix.

In another aspect, this invention provides a method for engineering a functional engineered fluorescent protein having a fluorescent property different than *Aequorea* green fluorescent protein, comprising substituting an amino acid that is located no more than 0.5 nm from any atom in the chromophore of an *Aequorea*-related green fluorescent protein with another amino acid; whereby the substitution alters a fluorescent property of the protein. In one embodiment, the amino acid substitution alters the electronic environment of the chromophore.

In another aspect, this invention provides a method for engineering a functional engineered fluorescent protein having a different fluorescent property than *Aequorea* green fluorescent protein comprising substituting amino acids in a loop domain of an *Aequorea*-related green fluorescent protein with amino acids so as to create a consensus sequence for phosphorylation or for proteolysis.

In another aspect, this invention provides a method for producing fluorescence resonance energy transfer comprising providing a donor molecule comprising a functional engineered fluorescent protein this invention; providing an appropriate acceptor molecule for the fluorescent protein; and bringing the donor molecule and

the acceptor molecule into sufficiently close contact to allow fluorescence resonance energy transfer.

In another aspect, this invention provides a method for producing fluorescence resonance energy transfer comprising providing an acceptor molecule comprising a functional engineered fluorescent protein of this invention; providing an appropriate donor molecule for the fluorescent protein; and bringing the donor molecule and the acceptor molecule into sufficiently close contact to allow fluorescence resonance energy transfer. In one embodiment, the donor molecule is a engineered fluorescent protein whose amino acid sequence comprises the substitution T203I and the acceptor molecule is an engineered fluorescent protein whose amino acid sequence comprises the substitution T203X, wherein X is an aromatic amino acid selected from H, Y, W or F, said functional engineered fluorescent protein having a different fluorescent property than *Aequorea* green fluorescent protein.

In another aspect, this invention provides a crystal of a protein comprising a fluorescent protein with an amino acid sequence substantially identical to SEQ ID NO: 2, wherein said crystal diffracts with at least a 2.0 to 3.0 angstrom resolution.

In another embodiment, this invention provides computational method of designing a fluorescent protein comprising determining from a three dimensional model of a crystallized fluorescent protein comprising a fluorescent protein with a bound ligand, at least one interacting amino acid of the fluorescent protein that interacts with at least one first chemical moiety of the ligand, and selecting at least one chemical modification of the first chemical moiety to produce a second chemical moiety with a structure to either decrease or increase an interaction between the interacting amino acid and the second chemical moiety compared to the interaction between the interacting amino acid and the first chemical moiety.

In another embodiment, this invention provides a computational method of modeling the three dimensional structure of a fluorescent protein comprising determining a three dimensional relationship between at least two atoms listed in the atomic coordinates of Figs. 5-1 to 5-28.

In another embodiment, this invention provides a device comprising a storage device and, stored in the device, at least 10 atomic coordinates selected from the

atomic coordinates listed in Figs. 5-1 to 5-28. In one embodiment, the storage device is a computer readable device that stores code that receives as input the atomic coordinates. In another embodiment, the computer readable device is a floppy disk or a hard drive.

In another embodiment this invention provides a nucleic acid molecule comprising a nucleotide sequence encoding a functional engineered fluorescent protein whose amino acid sequence is substantially identical to the amino acid sequence of *Aequorea* green fluorescent protein (SEQ ID NO:2) and which differs from SEQ ID NO:2 by at least one first substitution at position T203, wherein the substitution selected from the group consisting H, Y, W or F, and at least one second substitution at position H148.

In another aspect the present invention includes a method of determining the presence of an anion of interest in a sample, comprising the steps of introducing an engineered green fluorescent protein into a sample, said engineered green fluorescent protein comprising an amino acid sequence substantially identical to the amino acid sequence of *Aequorea* green fluorescent protein (SEQ ID NO:2) and which differs from SEQ ID NO:2 by at least one first substitution at position T203, wherein the substitution selected from the group consisting H, Y, W or F, and determining the fluorescence of said engineered green fluorescent protein in said sample.

In another embodiment, the invention includes a functional engineered fluorescent protein whose amino acid sequence is substantially identical to the amino acid sequence of *Aequorea* green fluorescent protein (SEQ ID NO:2) and which differs from SEQ ID NO:2 by at least one first substitution at position T203, wherein the substitution selected from the group consisting H, Y, W or F, and at least one second substitution at position H148, wherein said functional engineered fluorescent protein has a different fluorescent property than *Aequorea* green fluorescent protein.

In another embodiment the invention includes a host cell comprising a functional engineered fluorescent protein whose amino acid sequence is substantially identical to the amino acid sequence of *Aequorea* green fluorescent protein (SEQ ID NO:2) and which differs from SEQ ID NO:2 by at least one first substitution at position T203, wherein the substitution selected from the group consisting H, Y, W or

F, and at least one second substitution at position H148, wherein said functional engineered fluorescent protein has a different fluorescent property than *Aequorea* green fluorescent protein.

DETAILED DESCRIPTION OF THE INVENTION

I. DEFINITIONS

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by those of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described. For purposes of the present invention, the following terms are defined below.

"Binding pair" refers to two moieties (e.g. chemical or biochemical) that have an affinity for one another. Examples of binding pairs include antigen/antibodies, lectin/avidin, target polynucleotide/probe oligonucleotide, antibody/anti-antibody, receptor/ligand, enzyme/ligand and the like. "One member of a binding pair" refers to one moiety of the pair, such as an antigen or ligand.

"Nucleic acid" refers to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and, unless otherwise limited, encompasses known analogs of natural nucleotides that can function in a similar manner as naturally occurring nucleotides. It will be understood that when a nucleic acid molecule is represented by a DNA sequence, this also includes RNA molecules having the corresponding RNA sequence in which "U" replaces "T."

"Recombinant nucleic acid molecule" refers to a nucleic acid molecule which is not naturally occurring, and which comprises two nucleotide sequences which are not naturally joined together. Recombinant nucleic acid molecules are produced by artificial recombination, e.g., genetic engineering techniques or chemical synthesis.

Reference to a nucleotide sequence "encoding" a polypeptide means that the sequence, upon transcription and translation of mRNA, produces the polypeptide. This includes both the coding strand, whose nucleotide sequence is identical to mRNA and whose sequence is usually provided in the sequence listing, as well as its complementary strand, which is used as the template for transcription. As any person skilled in the art recognizes, this also includes all degenerate nucleotide sequences encoding the same amino acid sequence. Nucleotide sequences encoding a polypeptide include sequences containing introns.

"Expression control sequences" refers to nucleotide sequences that regulate the expression of a nucleotide sequence to which they are operatively linked. Expression control sequences are "operatively linked" to a nucleotide sequence when the expression control sequences control and regulate the transcription and, as appropriate, translation of the nucleotide sequence. Thus, expression control sequences can include appropriate promoters, enhancers, transcription terminators, a start codon (i.e., ATG) in front of a protein-encoding gene, splicing signals for introns, maintenance of the correct reading frame of that gene to permit proper translation of the mRNA, and stop codons.

"Naturally-occurring" as used herein, as applied to an object, refers to the fact that an object can be found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses) that can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory is naturally-occurring.

"Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences, such as when the appropriate molecules (e.g., inducers and polymerases) are bound to the control or regulatory sequence(s).

"Control sequence" refers to polynucleotide sequences which are necessary to effect the expression of coding and non-coding sequences to which they are ligated. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and transcription termination sequence; in eukaryotes, generally, such control sequences include promoters and transcription termination sequence. The term "control sequences" is intended to include, at a minimum, components whose presence can influence expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

"Isolated polynucleotide" refers a polynucleotide of genomic, cDNA, or synthetic origin or some combination there of, which by virtue of its origin the "isolated polynucleotide" (1) is not associated with the cell in which the "isolated polynucleotide" is found in nature, or (2) is operably linked to a polynucleotide which it is not linked to in nature.

"Polynucleotide" refers to a polymeric form of nucleotides of at least 10 bases in length, either ribonucleotides or deoxynucleotides or a modified form of either type of nucleotide. The term includes single and double stranded forms of DNA.

The term "probe" refers to a substance that specifically binds to another substance (a "target"). Probes include, for example, antibodies, nucleic acids, receptors and their ligands.

"Modulation" refers to the capacity to either enhance or inhibit a functional property of biological activity or process (e.g., enzyme activity or receptor binding); such enhancement or inhibition may be contingent on the occurrence of a specific event, such as activation of a signal transduction pathway, and/or may be manifest only in particular cell types.

The term "modulator" refers to a chemical (naturally occurring or non-naturally occurring), such as a synthetic molecule (e.g., nucleic acid, protein, non-peptide, or organic molecule), or an extract made from biological materials such as bacteria, plants, fungi, or animal (particularly mammalian) cells or tissues. Modulators can be evaluated for potential activity as inhibitors or activators (directly or indirectly) of a biological process or processes (e.g., agonist, partial antagonist, partial agonist, inverse agonist, antagonist, antineoplastic agents, cytotoxic agents, inhibitors of neoplastic transformation or cell proliferation, cell proliferation-promoting agents, and the like) by inclusion in screening assays described herein. The activity of a modulator may be known, unknown or partially known.

The term "test chemical" refers to a chemical to be tested by one or more screening method(s) of the invention as a putative modulator. A test chemical is usually not known to bind to the target of interest. The term "control test chemical" refers to a chemical known to bind to the target (e.g., a known agonist, antagonist,

partial agonist or inverse agonist). Usually, various predetermined concentrations of test chemicals are used for screening, such as .01 μM, .1 μM, 1.0 μM, and 10.0 μM.

The term "target" refers to a biochemical entity involved a biological process. Targets are typically proteins that play a useful role in the physiology or biology of an organism. A therapeutic chemical binds to target to alter or modulate its function. As used herein targets can include cell surface receptors, G-proteins, kinases, ion channels, phopholipases and other proteins mentioned herein.

The term "label" refers to a composition detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include ^{32}P, fluorescent dyes, fluorescent proteins, electron-dense reagents, enzymes (e.g., as commonly used in an ELISA), biotin, dioxigenin, or haptens and proteins for which antisera or monoclonal antibodies are available. For example, polypeptides of this invention can be made as detectable labels, by e.g., incorporating a them as into a polypeptide, and used to label antibodies specifically reactive with the polypeptide. A label often generates a measurable signal, such as radioactivity, fluorescent light or enzyme activity, which can be used to quantitate the amount of bound label.

The term "nucleic acid probe" refers to a nucleic acid molecule that binds to a specific sequence or sub-sequence of another nucleic acid molecule. A probe is preferably a nucleic acid molecule that binds through complementary base pairing to the full sequence or to a sub-sequence of a target nucleic acid. It will be understood that probes may bind target sequences lacking complete complementarity with the probe sequence depending upon the stringency of the hybridization conditions. Probes are preferably directly labelled as with isotopes, chromophores, lumiphores, chromogens, fluorescent proteins, or indirectly labelled such as with biotin to which a streptavidin complex may later bind. By assaying for the presence or absence of the probe, one can detect the presence or absence of the select sequence or sub-sequence.

A "labeled nucleic acid probe" is a nucleic acid probe that is bound, either covalently, through a linker, or through ionic, van der Waals or hydrogen bonds to a label such that the presence of the probe may be detected by detecting the presence of the label bound to the probe.

The terms "polypeptide" and "protein" refers to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical analogue of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers. The term "recombinant protein" refers to a protein that is produced by expression of a nucleotide sequence encoding the amino acid sequence of the protein from a recombinant DNA molecule.

The term "recombinant host cell" refers to a cell that comprises a recombinant nucleic acid molecule. Thus, for example, recombinant host cells can express genes that are not found within the native (non-recombinant) form of the cell.

The terms "isolated" "purified" or "biologically pure" refer to material which is substantially or essentially free from components which normally accompany it as found in its native state. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein or nucleic acid molecule which is the predominant protein or nucleic acid species present in a preparation is substantially purified. Generally, an isolated protein or nucleic acid molecule will comprise more than 80% of all macromolecular species present in the preparation. Preferably, the protein is purified to represent greater than 90% of all macromolecular species present. More preferably the protein is purified to greater than 95%, and most preferably the protein is purified to essential homogeneity, wherein other macromolecular species are not detected by conventional techniques.

The term "naturally-occurring" as applied to an object refers to the fact that an object can be found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses) that can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory is naturally-occurring.

The term "antibody" refers to a polypeptide substantially encoded by an immunoglobulin gene or immunoglobulin genes, or fragments thereof, which specifically bind and recognize an analyte (antigen). The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon and mu constant region

genes, as well as the myriad immunoglobulin variable region genes. Antibodies exist, e.g., as intact immunoglobulins or as a number of well characterized fragments produced by digestion with various peptidases. This includes, e.g., Fab' and $F(ab)'_2$ fragments. The term "antibody," as used herein, also includes antibody fragments either produced by the modification of whole antibodies or those synthesized *de novo* using recombinant DNA methodologies.

The term "immunoassay" refers to an assay that utilizes an antibody to specifically bind an analyte. The immunoassay is characterized by the use of specific binding properties of a particular antibody to isolate, target, and/or quantify the analyte.

The term "identical" in the context of two nucleic acid or polypeptide sequences refers to the residues in the two sequences which are the same when aligned for maximum correspondence. When percentage of sequence identity is used in reference to proteins or peptides it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acids residues are substituted for other amino acid residues with similar chemical properties (e.g. charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Where sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, *e.g.*, according to known algorithm. *See, e.g.,* Meyers and Miller, Computer Applic. Biol. Sci., 4: 11-17 (1988); Smith and Waterman (1981) *Adv. Appl. Math.* 2: 482; Needleman and Wunsch (1970) *J. Mol. Biol.* 48: 443; Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. USA* 85: 2444; Higgins and Sharp (1988) *Gene*, 73: 237-244 and Higgins and Sharp (1989) *CABIOS* 5: 151-153; Corpet, *et al.* (1988) *Nucleic Acids Research* 16, 10881-90; Huang, *et al.* (1992) *Computer Applications in the*

Biosciences 8, 155-65, and Pearson, *et al.* (1994) *Methods in Molecular Biology* 24, 307-31. Alignment is also often performed by inspection and manual alignment.

"Conservatively modified variations" of a particular nucleic acid sequence refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given polypeptide. For instance, the codons CGU, CGC, CGA, CGG, AGA, and AGG all encode the amino acid arginine. Thus, at every position where an arginine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of "conservatively modified variations." Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule by standard techniques. Accordingly, each "silent variation" of a nucleic acid which encodes a polypeptide is implicit in each described sequence. Furthermore, one of skill will recognize that individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids (typically less than 5%, more typically less than 1%) in an encoded sequence are "conservatively modified variations" where the alterations result in the substitution of an amino acid with a chemically similar amino acid. Conservative amino acid substitutions providing functionally similar amino acids are well known in the art. The following six groups each contain amino acids that are conservative substitutions for one another:

1) Alanine (A), Serine (S), Threonine (T);

2) Aspartic acid (D), Glutamic acid (E);

3) Asparagine (N), Glutamine (Q);

4) Arginine (R), Lysine (K);

5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and

6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

The term "complementary" means that one nucleic acid molecule has the sequence of the binding partner of another nucleic acid molecule. Thus, the sequence 5'-ATGC-3' is complementary to the sequence 5'-GCAT-3'.

An amino acid sequence or a nucleotide sequence is "substantially identical" or "substantially similar" to a reference sequence if the amino acid sequence or nucleotide sequence has at least 80% sequence identity with the reference sequence over a given comparison window. Thus, substantially similar sequences include those having, for example, at least 85% sequence identity, at least 90% sequence identity, at least 95% sequence identity or at least 99% sequence identity. Two sequences that are identical to each other are, of course, also substantially identical.

A subject nucleotide sequence is "substantially complementary" to a reference nucleotide sequence if the complement of the subject nucleotide sequence is substantially identical to the reference nucleotide sequence.

The term "stringent conditions" refers to a temperature and ionic conditions used in nucleic acid hybridization. Stringent conditions are sequence dependent and are different under different environmental parameters. Generally, stringent conditions are selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe.

The term "allelic variants" refers to polymorphic forms of a gene at a particular genetic locus, as well as cDNAs derived from mRNA transcripts of the genes and the polypeptides encoded by them.

The term "preferred mammalian codon" refers to the subset of codons from among the set of codons encoding an amino acid that are most frequently used in proteins expressed in mammalian cells as chosen from the following list:

Amino Acid	Preferred codons for high level mammalian expression
Gly	GGC,GGG
Glu	GAG
Asp	GAC
Val	GUG,GUC
Ala	GCC,GCU
Ser	AGC,UCC
Lys	AAG
Asn	AAC
Met	AUG
Ile	AUC
Thr	ACC
Trp	UGG
Cys	UGC
Tyr	UAU,UAC
Leu	CUG
Phe	UUC
Arg	CGC,AGG,AGA
Gln	CAG
His	CAC
Pro	CCC

Fluorescent molecules are useful in fluorescence resonance energy transfer ("FRET"). FRET involves a donor molecule and an acceptor molecule. To optimize the efficiency and detectability of FRET between a donor and acceptor molecule, several factors need to be balanced. The emission spectrum of the donor should overlap as much as possible with the excitation spectrum of the acceptor to maximize the overlap integral. Also, the quantum yield of the donor moiety and the extinction coefficient of the acceptor should likewise be as high as possible to maximize R_0, the distance at which energy transfer efficiency is 50%. However, the excitation spectra of the donor and acceptor should overlap as little as possible so that a wavelength

region can be found at which the donor can be excited efficiently without directly exciting the acceptor. Fluorescence arising from direct excitation of the acceptor is difficult to distinguish from fluorescence arising from FRET. Similarly, the emission spectra of the donor and acceptor should overlap as little as possible so that the two emissions can be clearly distinguished. High fluorescence quantum yield of the acceptor moiety is desirable if the emission from the acceptor is to be measured either as the sole readout or as part of an emission ratio. One factor to be considered in choosing the donor and acceptor pair is the efficiency of fluorescence resonance energy transfer between them. Preferably, the efficiency of FRET between the donor and acceptor is at least 10%, more preferably at least 50% and even more preferably at least 80%.

The term "fluorescent property" refers to the molar extinction coefficient at an appropriate excitation wavelength, the fluorescence quantum efficiency, the shape of the excitation spectrum or emission spectrum, the excitation wavelength maximum and emission wavelength maximum, the ratio of excitation amplitudes at two different wavelengths, the ratio of emission amplitudes at two different wavelengths, the excited state lifetime, or the fluorescence anisotropy. A measurable difference in any one of these properties between wild-type *Aequorea* GFP and the mutant form is useful. A measurable difference can be determined by determining the amount of any quantitative fluorescent property, e.g., the amount of fluorescence at a particular wavelength, or the integral of fluorescence over the emission spectrum. Determining ratios of excitation amplitude or emission amplitude at two different wavelengths ("excitation amplitude ratioing" and "emission amplitude ratioing", respectively) are particularly advantageous because the ratioing process provides an internal reference and cancels out variations in the absolute brightness of the excitation source, the sensitivity of the detector, and light scattering or quenching by the sample.

II. LONG WAVELENGTH ENGINEERED FLUORESCENT PROTEINS

A. Fluorescent Proteins

As used herein, the term "fluorescent protein" refers to any protein capable of fluorescence when excited with appropriate electromagnetic radiation. This includes fluorescent proteins whose amino acid sequences are either naturally occurring or engineered (i.e., analogs or mutants). Many cnidarians use green fluorescent proteins ("GFPs") as energy-transfer acceptors in bioluminescence. A "green fluorescent protein," as used herein, is a protein that fluoresces green light. Similarly, "blue fluorescent proteins" fluoresce blue light and "red fluorescent proteins" fluoresce red light. GFPs have been isolated from the Pacific Northwest jellyfish, *Aequorea victoria*, the sea pansy, *Renilla reniformis*, and *Phialidium gregarium*. W.W. Ward et al., *Photochem. Photobiol.*, 35:803-808 (1982); L.D. Levine et al., *Comp. Biochem. Physiol.*, 72B:77-85 (1982).

A variety of *Aequorea*-related fluorescent proteins having useful excitation and emission spectra have been engineered by modifying the amino acid sequence of a naturally occurring GFP from *Aequorea victoria*. (D.C. Prasher et al., *Gene*, 111:229-233 (1992); R. Heim et al., *Proc. Natl. Acad. Sci., USA*, 91:12501-04 (1994); U.S. patent application 08/337,915, filed November 10, 1994; International application PCT/US95/14692, filed 11/10/95.)

As used herein, a fluorescent protein is an "*Aequorea*-related fluorescent protein" if any contiguous sequence of 150 amino acids of the fluorescent protein has at least 85% sequence identity with an amino acid sequence, either contiguous or non-contiguous, from the 238 amino-acid wild-type *Aequorea* green fluorescent protein of Fig. 3 (SEQ ID NO:2). More preferably, a fluorescent protein is an *Aequorea*-related fluorescent protein if any contiguous sequence of 200 amino acids of the fluorescent protein has at least 95% sequence identity with an amino acid sequence, either contiguous or non-contiguous, from the wild type *Aequorea* green fluorescent protein of Fig. 3 (SEQ ID NO:2). Similarly, the fluorescent protein may be related to *Renilla* or *Phialidium* wild-type fluorescent proteins using the same standards.

Aequorea-related fluorescent proteins include, for example and without limitation, wild-type (native) *Aequorea victoria* GFP (D.C. Prasher et al., "Primary

structure of the *Aequorea victoria* green fluorescent protein," *Gene*, (1992) 111:229-33), whose nucleotide sequence (SEQ ID NO:1) and deduced amino acid sequence (SEQ ID NO:2) are presented in Fig. 3; allelic variants of this sequence, e.g., Q80R, which has the glutamine residue at position 80 substituted with arginine (M. Chalfie et al., *Science*, (1994) 263:802-805); those engineered *Aequorea*-related fluorescent proteins described herein, e.g., in Table A or Table F, variants that include one or more folding mutations and fragments of these proteins that are fluorescent, such as *Aequorea* green fluorescent protein from which the two amino-terminal amino acids have been removed. Several of these contain different aromatic amino acids within the central chromophore and fluoresce at a distinctly shorter wavelength than wild type species. For example, engineered proteins P4 and P4-3 contain (in addition to other mutations) the substitution Y66H, whereas W2 and W7 contain (in addition to other mutations) Y66W. Other mutations both close to the chromophore region of the protein and remote from it in primary sequence may affect the spectral properties of GFP and are listed in the first part of the table below.

TABLE A

Clone	Mutation(s)	Excitation max (nm)	Emission max (nm)	Extinct. Coeff. ($M^{-1}cm^{-1}$)	Quantum yield
Wild type	None	395 (475)	508	21,000 (7,150)	0.77
P4	Y66H	383	447	13,500	0.21
P4-3	Y66H Y145F	381	445	14,000	0.38
P4-3E	Y66H Y145F V163A	384	448	22,000	0.27
W7	Y66W N146I M153T V163A N212K	433 (453)	475 (501)	18,000 (17,100)	0.67
W2	Y66W I123V Y145H H148R M153T V163A N212K	432 (453)	480	10,000 (9,600)	0.72

W1C	S65A Y66W S72A N146I M153T V163A	435	495	21,200	0.39
W1B	F64L S65T Y66W N146I M153T V163A	434 (452)	476 (505)	32,500	0.4
S65T	S65T	489	511	39,200	0.68
P4-1	S65T M153A K238E	504 (396)	514	14,500 (8,600)	0.53
Emerald	S65T, S72A, N149K, M153T, I167T	487	509	57,500	0.68
EGFP	F64L, S65T	488	507	55,900	0.64
S65A	S65A	471	504		
S65C	S65C	479	507		
S65L	S65L	484	510		
Y66F	Y66F	360	442		
Y66W	Y66W	458	480		
Topaz	S65G S72A K79R T203Y	514	527	94,500	0.6
10C YFP	S65G V68L S72A T203Y	514	527	83,400	0.61
Sapphire	S72A, Y145F T203I	399	511	29,000	0.64

Additional mutations in *Aequorea*-related fluorescent proteins, referred to as "folding mutations," improve the ability of fluorescent proteins to fold at higher temperatures, and to be more fluorescent when expressed in mammalian cells, but have little or no effect on the peak wavelengths of excitation and emission. It should be noted that these may be combined with mutations that influence the spectral properties of GFP to produce proteins with altered spectral and folding properties. Folding mutations include: F64L, V68L, S72A, and also T44A, F99S, Y145F, N146I, M153T or A, V163A, I167T, S175G, S205T and N212K.

As used herein, the term "loop domain" refers to an amino acid sequence of an *Aequorea*-related fluorescent protein that connects the amino acids involved in the secondary structure of the eleven strands of the -barrel or the central -helix (residues 56-72) (see Fig. 1A and 1B).

As used herein, the "fluorescent protein moiety" of a fluorescent protein is that portion of the amino acid sequence of a fluorescent protein which, when the amino acid sequence of the fluorescent protein substrate is optimally aligned with the amino acid sequence of a naturally occurring fluorescent protein, lies between the amino terminal and carboxy terminal amino acids, inclusive, of the amino acid sequence of the naturally occurring fluorescent protein.

It has been found that fluorescent proteins can be genetically fused to other target proteins and used as markers to identify the location and amount of the target protein produced. Accordingly, this invention provides fusion proteins comprising a fluorescent protein moiety and additional amino acid sequences. Such sequences can be, for example, up to about 15, up to about 50, up to about 150 or up to about 1000 amino acids long. The fusion proteins possess the ability to fluoresce when excited by electromagnetic radiation. In one embodiment, the fusion protein comprises a polyhistidine tag to aid in purification of the protein.

B. Use Of The Crystal Structure Of Green Fluorescent Protein To Design Mutants Having Altered Fluorescent Characteristics

Using X-ray crystallography and computer processing, we have created a model of the crystal structure of *Aequorea* green fluorescent protein showing the relative location of the atoms in the molecule. This information is useful in identifying amino acids whose substitution alters fluorescent properties of the protein.

Fluorescent characteristics of *Aequorea*-related fluorescent proteins depend, in part, on the electronic environment of the chromophore. In general, amino acids that are within about 0.5 nm of the chromophore influence the electronic environment of the chromophore. Therefore, substitution of such amino acids can produce fluorescent proteins with altered fluorescent characteristics. In the excited state, electron density tends to shift from the phenolate towards the carbonyl end of the chromophore. Therefore, placement of increasing positive charge near the carbonyl end of the chromophore tends to decrease the energy of the excited state and cause a red-shift in the absorbance and emission wavelength maximum of the protein. Decreasing positive charge near the carbonyl end of the chromophore tends to have the opposite effect, causing a blue-shift in the protein's wavelengths.

Amino acids with charged (ionized D, E, K, and R), dipolar (H, N, Q, S, T, and uncharged D, E and K), and polarizable side groups (e.g., C, F, H, M, W and Y) are useful for altering the electronic environment of the chromophore, especially when they substitute an amino acid with an uncharged, nonpolar or non-polarizable side chain. In general, amino acids with polarizable side groups alter the electronic environment least, and, consequently, are expected to cause a comparatively smaller change in a fluorescent property. Amino acids with charged side groups alter the environment most, and, consequently, are expected to cause a comparatively larger change in a fluorescent property. However, amino acids with charged side groups are more likely to disrupt the structure of the protein and to prevent proper folding if buried next to the chromophore without any additional solvation or salt bridging. Therefore charged amino acids are most likely to be tolerated and to give useful effects when they replace other charged or highly polar amino acids that are already

solvated or involved in salt bridges. In certain cases, where substitution with a polarizable amino acid is chosen, the structure of the protein may make selection of a larger amino acid, e.g., W, less appropriate. Alternatively, positions occupied by amino acids with charged or polar side groups that are unfavorably oriented may be substituted with amino acids that have less charged or polar side groups. In another alternative, an amino acid whose side group has a dipole oriented in one direction in the protein can be substituted with an amino acid having a dipole oriented in a different direction.

More particularly, Table B lists several amino acids located within about 0.5 nm from the chromophore whose substitution can result in altered fluorescent characteristics. The table indicates, underlined, preferred amino acid substitutions at the indicated location to alter a fluorescent characteristic of the protein. In order to introduce such substitutions, the table also provides codons for primers used in site-directed mutagenesis involving amplification. These primers have been selected to encode economically the preferred amino acids, but they encode other amino acids as well, as indicated, or even a stop codon, denoted by Z. In introducing substitutions using such degenerate primers the most efficient strategy is to screen the collection to identify mutants with the desired properties and then sequence their DNA to find out which of the possible substitutions is responsible. Codons are shown in double-stranded form with sense strand above, antisense strand below. In nucleic acid sequences, R=(A or g); Y=(C or T); M=(A or C); K=(g or T); S=(g or C); W=(A or T); H=(A, T, or C); B=(g, T, or C); V=(g, A, or C); D=(g, A, or T); N=(A, C, g, or T).

TABLE B

Original position and presumed role		Change to	Codon
L42	Aliphatic residue near C=N of chromophore	C<u>FH</u>LQR<u>WY</u>Z	5'YDS 3' 3'RHS 5'
V61	Aliphatic residue near central -CH= of chromophore	<u>FYHC</u>LR	YDC RHg

	T62	Almost directly above center of chromophore bridge	AVFS	KYC MRg
			DEHKNQ	VAS BTS
			FYHCLR	YDC RHg
	V68	Aliphatic residue near carbonyl and G67	FYHL	YWC RWg
	N121	Near C-N site of ring closure between T65 and G67	CFHLQRWYZ	YDS RHS
	Y145	Packs near tyrosine ring of chromophore	WCFL	TKS AMS
			DEHNKQ	VAS BTS
	H148	H-bonds to phenolate oxygen	FYNI	WWC WWg
			KQR	MRg KYC
	V150	Aliphatic residue near tyrosine ring of chromophore	FYHL	YWC RWg
	F165	Packs near tyrosine ring	CHQRWYZ	YRS RYS
	I167	Aliphatic residue near phenolate; I167T has effects	FYHL	YWC RWg
	T203	H-bonds to phenolic oxygen of chromophore	FHLQRWYZ	YDS RHS
	E222	Protonation regulates ionization of chromophore	HKNQ	MAS KTS

Examples of amino acids with polar side groups that can be substituted with polarizable side groups include, for example, those in Table C.

TABLE C

Original position and presumed role			Change to	Codon
	Q69	Terminates chain of H-bonding waters	KREG	RRg YYC
	Q94	H-bonds to carbonyl terminus of chromophore	DEHKNQ	VAS BTS
	Q183	Bridges Arg96 and center of chromophore bridge	HY	YAC RTG
			EK	RAg YTC
	N185	Part of H-bond network near carbonyl of chromophore	DEHNKQ	VAS BTS

In another embodiment, an amino acid that is close to a second amino acid within about 0.5 nm of the chromophore can, upon substitution, alter the electronic properties of the second amino acid, in turn altering the electronic environment of the chromophore. Table D presents two such amino acids. The amino acids, L220 and V224, are close to E222 and oriented in the same direction in the ß pleated sheet.

TABLE D

Original position and presumed role			Change to	Codon
	L220	Packs next to Glu222; to make GFP pH sensitive	HKNPQT	MMS KKS
	V224	Packs next to Glu222; to make GFP pH sensitive	HKNPQT	MMS KKS
			CFHLQRWYZ	YDS RHS

One embodiment of the invention includes a nucleic acid molecule comprising a nucleotide sequence encoding a functional engineered fluorescent protein whose amino acid sequence is substantially identical to the amino acid sequence of *Aequorea* green fluorescent protein (SEQ ID NO:2) and which differs from SEQ ID NO:2 by at least a substitution at Q69, wherein the functional engineered fluorescent protein has a different fluorescent property than *Aequorea* green fluorescent protein. Preferably, the substitution at Q69 is selected from the group of K, R, E and G. The Q69 substitution can be combined with other mutations to improve the properties of the protein, such as a functional mutation at S65.

One embodiment of the invention includes a nucleic acid molecule comprising a nucleotide sequence encoding a functional engineered fluorescent protein whose amino acid sequence is substantially identical to the amino acid sequence of *Aequorea* green fluorescent protein (SEQ ID NO:2) and which differs from SEQ ID NO:2 by at least a substitution at E222, but not including E222G, wherein the functional engineered fluorescent protein has a different fluorescent property than *Aequorea* green fluorescent protein. Preferably, the substitution at E222 is selected from the group of N and Q. The E222 substitution can be combined with other mutations to improve the properties of the protein, such as a functional mutation at F64.

One embodiment of the invention includes a nucleic acid molecule comprising a nucleotide sequence encoding a functional engineered fluorescent protein whose amino acid sequence is substantially identical to the amino acid sequence of *Aequorea* green fluorescent protein (SEQ ID NO:2) and which differs from SEQ ID NO:2 by at least a substitution at Y145, wherein the functional engineered fluorescent protein has a different fluorescent property than *Aequorea* green fluorescent protein.

Preferably, the substitution at Y145 is selected from the group of W, C, F, L, E, H, K and Q. The Y145 substitution can be combined with other mutations to improve the properties of the protein, such as a Y66.

The invention also includes computer related embodiments, including computational methods of using the crystal coordinates for designing new fluorescent protein mutations and devices for storing the crystal data, including coordinates. For instance the invention includes a device comprising a storage device and, stored in the device, at least 10 atomic coordinates selected from the atomic coordinates listed in Figs. 5-1 to 5-28. More coordinates can be storage depending of the complexity of the calculations or the objective of using the coordinates (e.g. about 100, 1,000, or more coordinates). For example, larger numbers of coordinates will be desirable for more detailed representations of fluorescent protein structure. Typically, the storage device is a computer readable device that stores code that it receives as input the atomic coordinates. Although, other storage means as known in the art are contemplated. The computer readable device can be a floppy disk or a hard drive.

C. Use Of The Crystal Structure Of YFP To Design Mutants Having Altered Anion Binding Characteristics

In another aspect the invention includes the use of X-ray crystallography and computer processing, to create a model of the crystal structure of YFP showing the relative location, and amino acids that interact with bound ions. This information is useful in identifying amino acids whose substitution alters the specificity and affinity of the binding site to various anions. Because the binding of the anion is close to the chromophore of YFP, binding results in a modulation of the fluorescent properties of YFP that can be used to monitor anion binding and therefore the concentration of the anion.

The anion binding site found in YFP-H148Q exhibits many of the characteristics generally found in halide binding sites in other proteins. In the case of the anion-containing cavity in YFP-H148Q, the binding site is amphiphilic in nature, with one side lined with polar and charged groups (Tyr203, the chromophore, Arg96, Gln69, and Gln183), and the other with hydrophobic residues (Ile152, Leu201, Val163, Val150, and Phe165).

The design of engineered fluorescent proteins with altered anion binding specificies requires consideration of a number of factors. For example, one of the most significant factors contributing to the anion affinity and selectivity is the electrostatic configuration and make up of the binding pocket. In YFP these include the groups listed in Table E below.

TABLE E

Original position	and presumed role
S65, Y66, G67	Forms chromophore, aromatic edge interaction with ion
Q69	Hydrogen bonds to ion
R96	Charge interaction with ion
Q183	Charge interaction with ion
Y203	Hydrogen bonds to ion (Y)

In general, anion binding can be improved by creating more and or tighter binding interactions between the anion of interest and polar groups within the binding pocket. For example either directly substituting the polar residues above with more polar residues, or by substituting residues of different sizes, that may interact more effectively with the anion, can improve ion binding. For example the size and position of the chromophore may be altered by the substitution of S65 to G, A, C, V, L, I or T; Y66 may be altered by substitution to H, F or W; Q69 may be substituted to N or K; R96 to K; Q183 to N or K.

Hydration energy

Additionally the binding of an anion in a buried cavity near the chromophore requires replacement of ion-solvent interactions with ion-protein interactions. Relative binding energies of monovalent anions to YFP (Table J) and YFP-H148Q (Jayaraman *et al.*, 2000) in relation to their hydration energy indicate that hydration forces make important contributions towards binding. In the following series of

monoanions, the hydration energies are ordered from weak to strong: $CO_4^- < I^- < NO_3^- < SCN^- < Br^- < Cl^- < F^-$ (Wright & Diamond, 1977). Polyatomic monoanions and iodide have relatively weak hydration energies, whereas the other halides interact more strongly with water. In case of the spherically symmetric halides, hydration energy increases with decreasing atomic volume (Born, 1920), which is why larger halides are easier to bury in the more hydrophobic environment of a protein's interior. The trend observed for anion binding to the YFPs roughly follows the above series (Table J). Protein interaction generally increases with decreasing hydration energy, with the exception of fluoride, which may not completely dehydrate upon protein binding due to its small size.

The development of higher affinity anion binding sites therefore requires the creation of sufficient ion-protein interactions for example by the substitution of hydrophobic residues that line the ion binding pocket with more polar residues with more hydrogen bonding potential. Examples for these type of substitutions for improving the ion binding for larger and smaller anions are presented in Table F

TABLE F

i) Mutation of amino acids around the ion binding pocket to increase binding affinity for smaller anions than iodide.

Original position and presumed role		Change to
V150	Lines binding pocket	S, T, Q, N
I152	Lines binding pocket	L, V, F, S, T, Q, N
V163	Lines binding pocket	S, T, Q, N
F165	Lines binding pocket	Y, W
H181	Lines binding pocket	F, W
Q183	Lines binding pocket	K, R, N
L201	Lines binding pocket	S, T, Q, N, V, I

ii) Mutation of amino acids around the ion binding pocket to increase binding affinity for larger anions than iodide.

V150	Lines binding pocket	A, C, M, G, S, L
I152	Lines binding pocket	A, C, M, G, S
V163	Lines binding pocket	A, C, M, G, S, L
F165	Lines binding pocket	Y, L
H181	Lines binding pocket	K, R
Q183	Lines binding pocket	N, S, C
L201	Lines binding pocket	A, C, M, G, S

Size of the binding pocket

The size and shape of the binding pocket may also be of particular importance due to the buried nature of the binding site for larger anions. TCA, with a mean geometric diameter of 6.2 Å (Halm & Frizzell, 1992), is apparently too large to interact with YFP to a measurable extent, whereas the somewhat smaller TFA does show weak binding (Table J). Improvements in the binding affinity of larger anions could thus be achieved via the substitution of amino acids lining the binding pocket with smaller residues, as outlined in Table F above, as well as increasing solvent accessibility as discussed below.

Conformational changes upon anion binding

A series of conformational changes of various side chains lining the binding pocket in YFP are necessary for halide binding. The largest movements are observed for Gln69, and Gln183 although the apolar side chains of Leu201, Ile152, Val150, and Val163 (Fig. 11) all undergo movements to increase the cavity size in the presence of a bound halide. Another approach towards tighter anion binding therefore is the

substitution of the residues that undergoe the most dramatic conformational change upon binding, for smaller residues. These changes may reduce the need for structural rearrangements upon binding thereby making anion binding, more energetically favorable. These changes include those listed in Table F above as well as the substitution of Q69 for N.

Solvent Accessibility

The results from the structural determinations of various mutations at His148 suggests that specific mutations at this position can result in overall structural adjustments in the beta barrel that can directly affect both solvent accessibility and the volume of the binding pocket. Substitution of His148 for example to smaller amino acids such Q, N, G, A, L, V and I would therefore be predicted to increase solvent access to the chromophore and therefore improve binding of larger anions. Likewise substitution of His 148 with larger amino acids such as F or W would be likely to reduce anion access to the chromophore. Similarly more subtle changes could be achieved by substituting positions 147 and 149 with smaller or larger amino acids.

These mutations will typically be introduced in the YFP template protein via oligo-mediated site directed mutagenesis to create libraries of mutant proteins that typically have a 10 % probability of containing the wild-type amino acid residue and a 90% probability of containing one of the various mutant residues. Using this approach it is possible to rapidly screen libraries containing various combinations of mutants to identify the best combinations for a specific anion of interest. Typically this process can be repeated iteratively to ensure that sequence space around the binding pocket has been completely explored for any specific anion of interest.

D. Production Of Engineered Fluorescent Proteins

Recombinant production of a fluorescent protein involves expressing a nucleic acid molecule having sequences that encode the protein.

In one embodiment, the nucleic acid encodes a fusion protein in which a single polypeptide includes the fluorescent protein moiety within a longer polypeptide. The longer polypeptide can include a second functional protein, such as FRET partner or a protein having a second function (e.g., an enzyme, antibody or other binding protein). Nucleic acids that encode fluorescent proteins are useful as starting materials.

The fluorescent proteins can be produced as fusion proteins by recombinant DNA technology. Recombinant production of fluorescent proteins involves expressing nucleic acids having sequences that encode the proteins. Nucleic acids encoding fluorescent proteins can be obtained by methods known in the art. Fluorescent proteins can be made by site-specific mutagenesis of other nucleic acids encoding fluorescent proteins, or by random mutagenesis caused by increasing the error rate of PCR of the original polynucleotide with 0.1 mM $MnCl_2$ and unbalanced nucleotide concentrations. See, *e.g.*, U.S. patent application 08/337,915, filed November 10, 1994 or International application PCT/US95/14692, filed 11/10/95. The nucleic acid encoding a green fluorescent protein can be isolated by polymerase chain reaction of cDNA from *A. victoria* using primers based on the DNA sequence of *A. victoria* green fluorescent protein, as presented in Fig. 3. PCR methods are described in, for example, U.S. Pat. No. 4,683,195; Mullis et al. (1987) *Cold Spring Harbor Symp. Quant. Biol.* 51:263; and Erlich, ed., *PCR Technology*, (Stockton Press, NY, 1989).

The construction of expression vectors and the expression of genes in transfected cells involves the use of molecular cloning techniques also well known in the art. Sambrook et al., *Molecular Cloning -- A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (1989) and *Current Protocols in Molecular Biology*, F.M. Ausubel et al., eds., (Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc.). The expression vector can be adapted for function in prokaryotes or eukaryotes by inclusion of appropriate promoters, replication sequences, markers, etc.

Nucleic acids used to transfect cells with sequences coding for expression of the polypeptide of interest generally will be in the form of an expression vector including expression control sequences operatively linked to a nucleotide sequence coding for expression of the polypeptide. As used, the term "nucleotide sequence coding for expression of" a polypeptide refers to a sequence that, upon transcription and translation of mRNA, produces the polypeptide. This can include sequences containing, *e.g.*, introns. Expression control sequences are operatively linked to a nucleic acid sequence when the expression control sequences control and regulate the transcription and, as appropriate, translation of the nucleic acid sequence. Thus, expression control sequences can include appropriate promoters, enhancers, transcription terminators, a start codon (*i.e.*, ATG) in front of a protein-encoding gene, splicing signals for introns, maintenance of the correct reading frame of that gene to permit proper translation of the mRNA, and stop codons.

Methods which are well known to those skilled in the art can be used to construct expression vectors containing the fluorescent protein coding sequence and appropriate transcriptional/translational control signals. These methods include *in vitro* recombinant DNA techniques, synthetic techniques and *in vivo* recombination/genetic recombination. (See, for example, the techniques described in Maniatis, *et al., Molecular Cloning A Laboratory Manual*, Cold Spring Harbor Laboratory, N.Y., 1989).

Transformation of a host cell with recombinant DNA may be carried out by conventional techniques as are well known to those skilled in the art. Where the host is prokaryotic, such as *E. coli*, competent cells which are capable of DNA uptake can be prepared from cells harvested after exponential growth phase and subsequently treated by the $CaCl_2$ method by procedures well known in the art. Alternatively, $MgCl_2$ or RbCl can be used. Transformation can also be performed after forming a protoplast of the host cell or by electroporation.

When the host is a eukaryote, such methods of transfection of DNA as calcium phosphate co-precipitates, conventional mechanical procedures such as microinjection, electroporation, insertion of a plasmid encased in liposomes, or virus vectors may be used. Eukaryotic cells can also be co-transfected with DNA

sequences encoding the fusion polypeptide of the invention, and a second foreign DNA molecule encoding a selectable phenotype, such as the herpes simplex thymidine kinase gene. Another method is to use a eukaryotic viral vector, such as simian virus 40 (SV40) or bovine papilloma virus, to transiently infect or transform eukaryotic cells and express the protein. (*Eukaryotic Viral Vectors*, Cold Spring Harbor Laboratory, Gluzman ed., 1982). Preferably, a eukaryotic host is utilized as the host cell as described herein.

Techniques for the isolation and purification of either microbially or eukaryotically expressed polypeptides of the invention may be by any conventional means such as, for example, preparative chromatographic separations and immunological separations such as those involving the use of monoclonal or polyclonal antibodies or antigen. In one embodiment recombinant fluorescent proteins can be produced by expression of nucleic acid encoding for the protein in *E. coli*. *Aequorea*-related fluorescent proteins are best expressed by cells cultured between about 15°C and 30°C but higher temperatures (e.g. 37°C) are possible. After synthesis, these enzymes are stable at higher temperatures (e.g., 37°C) and can be used in assays at those temperatures.

A variety of host-expression vector systems may be utilized to express fluorescent protein coding sequence. These include but are not limited to microorganisms such as bacteria transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA expression vectors containing a fluorescent protein coding sequence; yeast transformed with recombinant yeast expression vectors containing the fluorescent protein coding sequence; plant cell systems infected with recombinant virus expression vectors (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (*e.g.*, Ti plasmid) containing a fluorescent protein coding sequence; insect cell systems infected with recombinant virus expression vectors (*e.g.*, baculovirus) containing a fluorescent protein coding sequence; or animal cell systems infected with recombinant virus expression vectors (*e.g.*, retroviruses, adenovirus, vaccinia virus) containing a fluorescent protein coding sequence, or transformed animal cell systems engineered for stable expression.

Depending on the host/vector system utilized, any of a number of suitable transcription and translation elements, including constitutive and inducible promoters, transcription enhancer elements, transcription terminators, *etc.* may be used in the expression vector (see, *e.g.*, Bitter, *et al.*, Methods in Enzymology 153:516-544, 1987). For example, when cloning in bacterial systems, inducible promoters such as pL of bacteriophage Σ, plac, ptrp, ptac (ptrp-lac hybrid promoter) and the like may be used. When cloning in mammalian cell systems, promoters derived from the genome of mammalian cells (*e.g.*, metallothionein promoter) or from mammalian viruses (*e.g.*, the retrovirus long terminal repeat; the adenovirus late promoter; the vaccinia virus 7.5K promoter) may be used. Promoters produced by recombinant DNA or synthetic techniques may also be used to provide for transcription of the inserted fluorescent protein coding sequence.

In bacterial systems a number of expression vectors may be advantageously selected depending upon the use intended for the fluorescent protein expressed. For example, when large quantities of the fluorescent protein are to be produced, vectors which direct the expression of high levels of fusion protein products that are readily purified may be desirable. Those which are engineered to contain a cleavage site to aid in recovering fluorescent protein are preferred.

In yeast, a number of vectors containing constitutive or inducible promoters may be used. For a review see, Current Protocols in Molecular Biology, Vol. 2, Ed. Ausubel, *et al.,* Greene Publish. Assoc. & Wiley Interscience, Ch. 13, 1988; Grant, *et al.*, Expression and Secretion Vectors for Yeast, in Methods in Enzymology, Eds. Wu & Grossman, 31987, Acad. Press, N.Y., Vol. 153, pp.516-544, 1987; Glover, DNA Cloning, Vol. II, IRL Press, Wash., D.C., Ch. 3, 1986; and Bitter, Heterologous Gene Expression in Yeast, Methods in Enzymology, Eds. Berger & Kimmel, Acad. Press, N.Y., Vol. 152, pp. 673-684, 1987; and The Molecular Biology of the Yeast Saccharomyces, Eds. Strathern *et al.*, Cold Spring Harbor Press, Vols. I and II, 1982. A constitutive yeast promoter such as ADH or LEU2 or an inducible promoter such as GAL may be used (Cloning in Yeast, Ch. 3, R. Rothstein In: DNA Cloning Vol.11, A Practical Approach, Ed. DM Glover, IRL Press, Wash., D.C., 1986). Alternatively,

vectors may be used which promote integration of foreign DNA sequences into the yeast chromosome.

In cases where plant expression vectors are used, the expression of a fluorescent protein coding sequence may be driven by any of a number of promoters. For example, viral promoters such as the 35S RNA and 19S RNA promoters of CaMV (Brisson, *et al., Nature* 310:511-514, 1984), or the coat protein promoter to TMV (Takamatsu, *et al.*, *EMBO J.* 6:307-311, 1987) may be used; alternatively, plant promoters such as the small subunit of RUBISCO (Coruzzi, *et al.*, 1984, *EMBO J.* 3:1671-1680; Broglie, *et al.*, *Science* 224:838-843, 1984); or heat shock promoters, *e.g.*, soybean hsp17.5-E or hsp17.3-B (Gurley, *et al.*, *Mol. Cell. Biol.* 6:559-565, 1986) may be used. These constructs can be introduced into plant cells using Ti plasmids, Ri plasmids, plant virus vectors, direct DNA transformation, microinjection, electroporation, *etc.* For reviews of such techniques see, for example, Weissbach & Weissbach, *Methods for Plant Molecular Biology,* Academic Press, NY, Section VIII, pp. 421-463, 1988; and Grierson & Corey, *Plant Molecular Biology*, 2d Ed., Blackie, London, Ch. 7-9, 1988.

An alternative expression system which could be used to express fluorescent protein is an insect system. In one such system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes. The virus grows in *Spodoptera frugiperda* cells. The fluorescent protein coding sequence may be cloned into non-essential regions (for example, the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (for example the polyhedrin promoter). Successful insertion of the fluorescent protein coding sequence will result in inactivation of the polyhedrin gene and production of non-occluded recombinant virus (*i.e.*, virus lacking the proteinaceous coat coded for by the polyhedrin gene). These recombinant viruses are then used to infect *Spodoptera frugiperda* cells in which the inserted gene is expressed, see Smith, *et al., J. Viol.* 46:584, 1983; Smith, U.S. Patent No. 4,215,051.

Eukaryotic systems, and preferably mammalian expression systems, allow for proper post-translational modifications of expressed mammalian proteins to occur. Eukaryotic cells which possess the cellular machinery for proper processing of the

primary transcript, glycosylation, phosphorylation, and, advantageously secretion of the gene product should be used as host cells for the expression of fluorescent protein. Such host cell lines may include but are not limited to CHO, VERO, BHK, HeLa, COS, MDCK, Jurkat, HEK-293, and WI38.

Mammalian cell systems which utilize recombinant viruses or viral elements to direct expression may be engineered. For example, when using adenovirus expression vectors, the fluorescent protein coding sequence may be ligated to an adenovirus transcription/translation control complex, *e.g.*, the late promoter and tripartite leader sequence. This chimeric gene may then be inserted in the adenovirus genome by *in vitro* or *in vivo* recombination. Insertion in a non-essential region of the viral genome (*e.g.*, region E1 or E3) will result in a recombinant virus that is viable and capable of expressing the fluorescent protein in infected hosts (*e.g.*, see Logan & Shenk, *Proc. Natl. Acad. Sci. USA*, 81: 3655-3659, 1984). Alternatively, the vaccinia virus 7.5K promoter may be used. (*e.g.*, see, Mackett, *et al.*, *Proc. Natl. Acad. Sci. USA,*79: 7415-7419, 1982; Mackett, *et al., J. Virol.* 49: 857-864, 1984; Panicali, *et al.*, *Proc. Natl. Acad. Sci. USA* 79: 4927-4931, 1982). Of particular interest are vectors based on bovine papilloma virus which have the ability to replicate as extrachromosomal elements (Sarver, *et al.*, *Mol. Cell. Biol.* 1: 486, 1981). Shortly after entry of this DNA into mouse cells, the plasmid replicates to about 100 to 200 copies per cell. Transcription of the inserted cDNA does not require integration of the plasmid into the host's chromosome, thereby yielding a high level of expression. These vectors can be used for stable expression by including a selectable marker in the plasmid, such as the *neo* gene. Alternatively, the retroviral genome can be modified for use as a vector capable of introducing and directing the expression of the fluorescent protein gene in host cells (Cone & Mulligan, *Proc. Natl. Acad. Sci. USA*, 81:6349-6353, 1984). High level expression may also be achieved using inducible promoters, including, but not limited to, the metallothionine IIA promoter and heat shock promoters.

The invention can also include a localization sequence, such as a nuclear localization sequence, an endoplasmic reticulum localization sequence, a peroxisome localization sequence, a mitochondrial localization sequence, or a localized protein.

Localization sequences can be targeting sequences which are described, for example, in "Protein Targeting", chapter 35 of Stryer, L., *Biochemistry* (4th ed.). W.H. Freeman, 1995. The localization sequence can also be a localized protein. Some important localization sequences include those targeting the nucleus (KKKRK), mitochondrion (amino terminal MLRTSSLFTRRVQPSLFRNILRLQST-), endoplasmic reticulum (KDEL at C-terminus, assuming a signal sequence present at N-terminus), peroxisome (SKF at C-terminus), prenylation or insertion into plasma membrane (CaaX, CC, CXC, or CCXX at C-terminus), cytoplasmic side of plasma membrane (fusion to SNAP-25), or the Golgi apparatus (fusion to furin).

For long-term, high-yield production of recombinant proteins, stable expression is preferred. Rather than using expression vectors which contain viral origins of replication, host cells can be transformed with the fluorescent protein cDNA controlled by appropriate expression control elements (*e.g.*, promoter, enhancer, sequences, transcription terminators, polyadenylation sites, *etc.*), and a selectable marker. The selectable marker in the recombinant plasmid confers resistance to the selection and allows cells to stably integrate the plasmid into their chromosomes and grow to form foci which in turn can be cloned and expanded into cell lines. For example, following the introduction of foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then are switched to a selective media. A number of selection systems may be used, including but not limited to the herpes simplex virus thymidine kinase (Wigler, *et al.*, *Cell,* 11: 223, 1977), hypoxanthine-guanine phosphoribosyltransferase (Szybalska & Szybalski, *Proc. Natl. Acad. Sci. USA*, 48:2026, 1962), and adenine phosphoribosyltransferase (Lowy, *et al.*, *Cell*, 22: 817, 1980) genes can be employed in tk^-, $hgprt^-$ or $aprt^-$ cells respectively. Also, antimetabolite resistance can be used as the basis of selection for dhfr, which confers resistance to methotrexate (Wigler, *et al.*, *Proc. Natl. Acad. Sci. USA*, 77: 3567, 1980; O'Hare, *et al.*, *Proc. Natl. Acad. Sci. USA,* 8: 1527, 1981); gpt, which confers resistance to mycophenolic acid (Mulligan & Berg, *Proc. Natl. Acad. Sci. USA*, 78: 2072, 1981; neo, which confers resistance to the aminoglycoside G-418 (Colberre-Garapin, *et al.*, *J. Mol. Biol.,* 150:1, 1981); and hygro, which confers resistance to hygromycin (Santerre, *et al.*, *Gene,* 30: 147, 1984) genes. Recently,

additional selectable genes have been described, namely trpB, which allows cells to utilize indole in place of tryptophan; hisD, which allows cells to utilize histinol in place of histidine (Hartman & Mulligan, *Proc. Natl. Acad. Sci. USA,* 85:8047, 1988); and ODC (ornithine decarboxylase) which confers resistance to the ornithine decarboxylase inhibitor, 2-(difluoromethyl)-DL-ornithine, DFMO (McConlogue L., In: *Current Communications in Molecular Biology,* Cold Spring Harbor Laboratory, ed., 1987).

DNA sequences encoding the fluorescence protein polypeptide of the invention can be expressed *in vitro* by DNA transfer into a suitable host cell. "Host cells" are cells in which a vector can be propagated and its DNA expressed. The term also includes any progeny of the subject host cell. It is understood that all progeny may not be identical to the parental cell since there may be mutations that occur during replication. However, such progeny are included when the term "host cell" is used. Methods of stable transfer, in other words when the foreign DNA is continuously maintained in the host, are known in the art.

The expression vector can be transfected into a host cell for expression of the recombinant nucleic acid. Host cells can be selected for high levels of expression in order to purify the fluorescent protein fusion protein. *E. coli* is useful for this purpose. Alternatively, the host cell can be a prokaryotic or eukaryotic cell selected to study the activity of an enzyme produced by the cell. In this case, the linker peptide is selected to include an amino acid sequence recognized by the protease. The cell can be, *e.g.*, a cultured cell or a cell *in vivo*.

A primary advantage of fluorescent protein fusion proteins is that they are prepared by normal protein biosynthesis, thus completely avoiding organic synthesis and the requirement for customized unnatural amino acid analogs. The constructs can be expressed in *E. coli* in large scale for *in vitro* assays. Purification from bacteria is simplified when the sequences include polyhistidine tags for one-step purification by nickel-chelate chromatography. Alternatively, the substrates can be expressed directly in a desired host cell for assays *in situ*.

In another embodiment, the invention provides a transgenic non-human animal that expresses a nucleic acid sequence which encodes the fluorescent protein.

The "non-human animals" of the invention comprise any non-human animal having nucleic acid sequence which encodes a fluorescent protein. Such non-human animals include vertebrates such as rodents, non-human primates, sheep, dog, cow, pig, amphibians, and reptiles. Preferred non-human animals are selected from the rodent family including rat and mouse, most preferably mouse. The "transgenic non-human animals" of the invention are produced by introducing "transgenes" into the germline of the non-human animal. Embryonal target cells at various developmental stages can be used to introduce transgenes. Different methods are used depending on the stage of development of the embryonic target cell. The zygote is the best target for micro-injection. In the mouse, the male pronucleus reaches the size of approximately 20 micrometers in diameter which allows reproducible injection of 1-2 pl of DNA solution. The use of zygotes as a target for gene transfer has a major advantage in that in most cases the injected DNA will be incorporated into the host gene before the first cleavage (Brinster *et al., Proc. Natl. Acad. Sci. USA* 82:4438-4442, 1985). As a consequence, all cells of the transgenic non-human animal will carry the incorporated transgene. This will in general also be reflected in the efficient transmission of the transgene to offspring of the founder since 50% of the germ cells will harbor the transgene. Microinjection of zygotes is the preferred method for incorporating transgenes in practicing the invention.

The term "transgenic" is used to describe an animal which includes exogenous genetic material within all of its cells. A "transgenic" animal can be produced by cross-breeding two chimeric animals which include exogenous genetic material within cells used in reproduction. Twenty-five percent of the resulting offspring will be transgenic *i.e.*, animals which include the exogenous genetic material within all of their cells in both alleles. 50% of the resulting animals will include the exogenous genetic material within one allele and 25% will include no exogenous genetic material.

Retroviral infection can also be used to introduce transgene into a non-human animal. The developing non-human embryo can be cultured *in vitro* to the blastocyst stage. During this time, the blastomeres can be targets for retro viral infection (Jaenich, R., *Proc. Natl. Acad. Sci USA* 73:1260-1264, 1976). Efficient infection of

the blastomeres is obtained by enzymatic treatment to remove the zona pellucida (Hogan, *et al.* (1986) in Manipulating the Mouse Embryo, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.). The viral vector system used to introduce the transgene is typically a replication-defective retro virus carrying the transgene (Jahner, *et al.*, *Proc. Natl. Acad. Sci. USA* 82:6927-6931, 1985; Van der Putten, *et al.*, *Proc. Natl. Acad. Sci USA* 82:6148-6152, 1985). Transfection is easily and efficiently obtained by culturing the blastomeres on a monolayer of virus-producing cells (Van der Putten, *supra*; Stewart, *et al.*, *EMBO J.* 6:383-388, 1987). Alternatively, infection can be performed at a later stage. Virus or virus-producing cells can be injected into the blastocoele (D. Jahner *et al.*, *Nature* 298:623-628, 1982). Most of the founders will be mosaic for the transgene since incorporation occurs only in a subset of the cells which formed the transgenic nonhuman animal. Further, the founder may contain various retro viral insertions of the transgene at different positions in the genome which generally will segregate in the offspring. In addition, it is also possible to introduce transgenes into the germ line, albeit with low efficiency, by intrauterine retro viral infection of the midgestation embryo (D. Jahner *et al.*, *supra*).

A third type of target cell for transgene introduction is the embryonal stem cell (ES). ES cells are obtained from pre-implantation embryos cultured *in vitro* and fused with embryos (M. J. Evans *et al. Nature* 292:154-156, 1981; M.O. Bradley *et al.*, *Nature* 309: 255-258, 1984; Gossler, *et al.*, *Proc. Natl. Acad. Sci USA* 83: 9065-9069, 1986; and Robertson *et al.*, *Nature* 322:445-448, 1986). Transgenes can be efficiently introduced into the ES cells by DNA transfection or by retro virus-mediated transduction. Such transformed ES cells can thereafter be combined with blastocysts from a nonhuman animal. The ES cells thereafter colonize the embryo and contribute to the germ line of the resulting chimeric animal. (For review see Jaenisch, R., *Science* 240: 1468-1474, 1988).

"Transformed" means a cell into which (or into an ancestor of which) has been introduced, by means of recombinant nucleic acid techniques, a heterologous nucleic acid molecule. "Heterologous" refers to a nucleic acid sequence that either originates

from another species or is modified from either its original form or the form primarily expressed in the cell.

"Transgene" means any piece of DNA which is inserted by artifice into a cell, and becomes part of the genome of the organism (*i.e.*, either stably integrated or as a stable extrachromosomal element) which develops from that cell. Such a transgene may include a gene which is partly or entirely heterologous (*i.e.*, foreign) to the transgenic organism, or may represent a gene homologous to an endogenous gene of the organism. Included within this definition is a transgene created by the providing of an RNA sequence which is transcribed into DNA and then incorporated into the genome. The transgenes of the invention include DNA sequences which encode which encodes the fluorescent protein which may be expressed in a transgenic non-human animal. The term "transgenic" as used herein additionally includes any organism whose genome has been altered by *in vitro* manipulation of the early embryo or fertilized egg or by any transgenic technology to induce a specific gene knockout. The term "gene knockout" as used herein, refers to the targeted disruption of a gene *in vivo* with complete loss of function that has been achieved by any transgenic technology familiar to those in the art. In one embodiment, transgenic animals having gene knockouts are those in which the target gene has been rendered nonfunctional by an insertion targeted to the gene to be rendered non-functional by homologous recombination. As used herein, the term "transgenic" includes any transgenic technology familiar to those in the art which can produce an organism carrying an introduced transgene or one in which an endogenous gene has been rendered non-functional or "knocked out."

III. USES OF ENGINEERED FLUORESCENT PROTEINS

The proteins of this invention are useful in any methods that employ fluorescent proteins.

The engineered fluorescent proteins of this invention are useful as fluorescent markers in the many ways fluorescent markers already are used. This includes, for example, coupling engineered fluorescent proteins to antibodies, nucleic acids or

other receptors for use in detection assays, such as immunoassays or hybridization assays.

The engineered fluorescent proteins of this invention are useful to track the movement of proteins in cells. In this embodiment, a nucleic acid molecule encoding the fluorescent protein is fused to a nucleic acid molecule encoding the protein of interest in an expression vector. Upon expression inside the cell, the protein of interest can be localized based on fluorescence. In another version, two proteins of interest are fused with two engineered fluorescent proteins having different fluorescent characteristics.

The engineered fluorescent proteins of this invention are useful in systems to detect induction of transcription. In certain embodiments, a nucleotide sequence encoding the engineered fluorescent protein is fused to expression control sequences of interest and the expression vector is transfected into a cell. Induction of the promoter can be measured by detecting the expression and/or quantity of fluorescence. Such constructs can be used to follow signaling pathways from receptor to promoter.

The engineered fluorescent proteins of this invention are useful in applications involving FRET. Such applications can detect events as a function of the movement of fluorescent donors and acceptor towards or away from each other. One or both of the donor/acceptor pair can be a fluorescent protein. A preferred donor and receptor pair for FRET based assays is a donor with a T203I mutation and an acceptor with the mutation T203X, wherein X is an aromatic amino acid-39, especially T203Y, T203W, or T203H. In a particularly useful pair the donor contains the following mutations: S72A, K79R, Y145F, M153A and T203I (with a excitation peak of 395 nm and an emission peak of 511 nm) and the acceptor contains the following mutations S65G, S72A, K79R, and T203Y. This particular pair provides a wide separation between the excitation and emission peaks of the donor and provides good overlap between the donor emission spectrum and the acceptor excitation spectrum. Other red-shifted mutants, such as those described herein, can also be used as the acceptor in such a pair.

In one aspect, FRET is used to detect the cleavage of a substrate having the donor and acceptor coupled to the substrate on opposite sides of the cleavage site. Upon cleavage of the substrate, the donor/acceptor pair physically separate, eliminating FRET. Assays involve contacting the substrate with a sample, and determining a qualitative or quantitative change in FRET. In one embodiment, the engineered fluorescent protein is used in a substrate for β-lactamase. Examples of such substrates are described in United States patent applications 08/407,544, filed March 20, 1995 and International Application PCT/US96/04059, filed March 20, 1996. In another embodiment, an engineered fluorescent protein donor/acceptor pair are part of a fusion protein coupled by a peptide having a proteolytic cleavage site. Such tandem fluorescent proteins are described in United States patent application 08/594,575, filed January 31, 1996.

In another aspect, FRET is used to detect changes in potential across a membrane. A donor and acceptor are placed on opposite sides of a membrane such that one translates across the membrane in response to a voltage change. This creates a measurable FRET. Such a method is described in United States patent application 08/481,977, filed June 7, 1995 and International Application PCT/US96/09652, filed June 6, 1996.

The engineered proteins of this invention are useful in the creation of biosensors for determining the concentrations of ions within samples and living cells and transgenic organisms. Upon binding of an ion to the fluorescent protein, a change in at least one measurable fluorescent property of the engineered fluorescent protein occurs that provides the basis for determining the presence of the ion of interest.

The engineered protein of this invention are useful in the creation of fluorescent substrates for protein kinases. Such substrates incorporate an amino acid sequence recognizable by protein kinases. Upon phosphorylation, the engineered fluorescent protein undergoes a change in a fluorescent property. Such substrates are useful in detecting and measuring protein kinase activity in a sample of a cell, upon transfection and expression of the substrate. Preferably, the kinase recognition site is placed within about 20 amino acids of a terminus of the engineered fluorescent protein. The kinase recognition site also can be placed in a loop domain of the

protein. (See, e.g. Figure 1B.) Methods for making fluorescent substrates for protein kinases are described in United States patent application 08/680,877, filed July 16, 1996.

A protease recognition site also can be introduced into a loop domain. Upon cleavage, fluorescent property changes in a measurable fashion.

The invention also includes a method of identifying a test chemical. Typically, the method includes contacting a test chemical a sample containing a biological entity labeled with a functional, engineered fluorescent protein or a polynucleotide encoding said functional, engineered fluorescent protein. By monitoring fluorescence (i.e. a fluorescent property) from the sample containing the functional engineered fluorescent protein it can be determined whether a test chemical is active. Controls can be included to insure the specificity of the signal. Such controls include measurements of a fluorescent property in the absence of the test chemical, in the presence of a chemical with an expected activity (e.g., a known modulator) or engineered controls (e.g., absence of engineered fluorescent protein, absence of engineered fluorescent protein polynucleotide or the absence of operably linkage of the engineered fluorescent protein).

The fluorescence in the presence of a test chemical can be greater or less than in the absence of said test chemical. For instance if the engineered fluorescent protein is used a reporter of gene expression, the test chemical may up or down regulate gene expression. For such types of screening, the polynucleotide encoding the functional, engineered fluorescent protein is operatively linked to a genomic polynucleotide or a re. Alternatively, the functional, engineered fluorescent protein is fused to second functional protein. This embodiment can be used to track localization of the second protein or to track protein-protein interactions using energy transfer.

IV. PROCEDURES

Fluorescence in a sample is measured using a fluorimeter. In general, excitation radiation from an excitation source having a first wavelength, passes through excitation optics. The excitation optics cause the excitation radiation to

excite the sample. In response, fluorescent proteins in the sample emit radiation which has a wavelength that is different from the excitation wavelength. Collection optics then collect the emission from the sample. The device can include a temperature controller to maintain the sample at a specific temperature while it is being scanned. According to one embodiment, a multi-axis translation stage moves a microtiter plate holding a plurality of samples in order to position different wells to be exposed. The multi-axis translation stage, temperature controller, auto-focusing feature, and electronics associated with imaging and data collection can be managed by an appropriately programmed digital computer. The computer also can transform the data collected during the assay into another format for presentation. This process can be miniaturized and automated to enable screening many thousands of compounds.

Methods of performing assays on fluorescent materials are well known in the art and are described in, e.g., Lakowicz, J.R., *Principles of Fluorescence Spectroscopy*, New York:Plenum Press (1983); Herman, B., Resonance energy transfer microscopy, in: *Fluorescence Microscopy of Living Cells in Culture, Part B, Methods in Cell Biology*, vol. 30, ed. Taylor, D.L. & Wang, Y.-L., San Diego: Academic Press (1989), pp. 219-243; Turro, N.J., *Modern Molecular Photochemistry*, Menlo Park: Benjamin/Cummings Publishing Col, Inc. (1978), pp. 296-361.

Mutagenesis and protein preparation

YFP variants and revertants were prepared using the PCR-based QuikChange™ Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA), according to the manufacturer's directions and using the YFP clone 10c as a template (Ormo *et al.*, 1996) . Mutations were verified by sequencing the entire gene, and all GFP variants were expressed and purified as described (Ormo *et al.*, 1996) .

Fluorescence measurements

Small aliquots of concentrated protein were diluted 25-fold into a series of buffers (20 mM MES pH 6.0, MES pH 6.5, PIPES 7.0, HEPES 7.5, and TAPS 8.0) with constant ionic strength. The buffers contained varying concentrations of either potassium chloride or potassium iodide, and the ionic strength was adjusted to 150 mM with potassium D-gluconate. Fluorescence measurements as a function of pH and halide concentration were carried out on a Hitachi F4500 fluorescence spectrophotometer at room temperature (λ_{ex} = 514 nm), scanning the emission between 520 and 550 nm three times at a rate of 60 nm/min. Maximum emission at 528 nm was averaged, and normalized with respect to fluorescence in the absence of halides.

Crystal growth and data collection

YFP-H148Q was concentrated to 15 mg/ml in 20 mM TRIS pH 7.9, and crystals were grown in hanging drops containing 5 µl protein and 5 µl mother liquor. The mother liquor contained 22% PEG 1550 at pH 5.5 in 100 mM sodium acetate and 90 mM $MgCl_2$. The rod-shaped crystals were approximately 0.04 mm across and up to 1.0 mm long, and grew within 1.5 to 2 years at 4 C. One crystal was soaked in synthetic mother liquor containing the above ingredients without $MgCl_2$ but 100 mM potassium iodide, and 20% ethylene glycol for cryo-protection (referred to as iodide soak). Another crystal was soaked in the above mother liquor containing 100 mM $MgCl_2$ and 20% ethylene glycol (referred to as chloride soak). Both soaks were carried out at pH 4.6 for 4 hours at room temperature, and data collection proceeded immediately thereafter. The crystals were flash-frozen, and X-ray diffraction data were collected at 100 K using a RAXIS-IIc image plate mounted on a Rigaku RUH3 rotating anode generator equipped with mirrors.

Structure determination of YFP-H148Q, and identification of iodide binding sites

The two data sets were processed with Denzo v1.9 and scaled using ScalePack (Otwinowski & Minor, 1997). The spacegroup is $P2_12_12_1$, with unit cell parameters a = 51.2, b = 62.8, and c = 68.7 Å for the iodide soak, and a = 51.7, b = 62.6, and c = 66.2 Å for the chloride soak. The crystals are nearly isomorphous to YFP-H148G (Wachter et al., 1998) and GFP S65T crystals (Ormo *et al.*, 1996) previously described, and the YFP-H148G coordinate file 2yfp (Wachter *et al.*, 1998) was used as a model for phasing. A model for the anionic chromophore was obtained by semi-empirical molecular orbital calculations using AM1 in the program SPARTAN version 4.1 (Wavefunction Inc., Irvine, CA).

An anomalous difference map was calculated from the data set derived from the iodide soak (anomalous data 65% complete), since iodine exhibits a significant anomalous signal at the in-house CuK_α wavelength of 1.54 Å. Heavy atom phases were approximated by subtraction of 90° from calculated protein phases using the program scaleit in the CCP4 program suite (Collaborative Computational Project N. 4, 1994). The anomalous difference map identified two iodide positions, one buried in the protein interior and one on the protein surface.

Refinement of YFP-H148Q with and without bound iodide

The two datasets, derived from the iodide and from the chloride soak, were refined in a similar manner. After initial rigid body refinement to 4.0 Å, positional refinement was carried out using the data to 3.0 Å, then to the limit of resolution (Table 1), using the program TNT (Tronrud *et al.*, 1987) . During early cycles of refinement, bound halides were not modeled, and the glutamine in position 148 was modeled as a glycine. Electron density maps ($2F_O - F_C$ and $F_O - F_C$) were inspected intermittently using O (Jones *et al.*, 1991) . The $F_O - F_C$ maps clearly indicated the positions of the buried and surface iodides, at 11 and 5.5 rms deviations respectively, located in the centers of the two anomalous difference density peaks, though no positive difference density consistent with buried chloride binding was observed. Density for the Gln148 side chain was clearly visible early on, allowing for the modeling of the glutamine as a rotamer different from the original histidine.

B-factors were refined using the default TNT B-factor correlation library. B-factor correlation values derived from His and Phe were used to model the chromophore atoms. Bound solvent molecules were added to the model where appropriate as judged from difference density and proximity of hydrogen bond partners. Before refining the occupancy of the bound halides, the B-factors for these halides were fixed. The thermal factor of the buried iodide was set to the average B-factor of the twelve atoms closest to it, 30 $Å^2$, (Figure 5), and the thermal factor for the surface iodide was set to the average B-factor of the six closest solvent molecules bound to the protein surface, 39 $Å^2$. The last step in refinement was the refinement of the occupancy of the two bound halides.

Determination of chromophore pK_a and iodide binding constants by absorbance

The chromophore pK_a was determined from absorbance scans at varying anion concentrations. Absorbance scans were collected at room temperature between 250 and 600 nm (Shimadzu 2101 spectrophotometer) on 0.05 mg/ml YFP under two different pH conditions appropriate for the particular anion, chosen from a series of buffers (20 mM malic acid pH 5.8, malic acid pH 6.1, MES pH 6.4, HEPES pH 7.1). The optical density of the chromophore anion (514 to 515 nm for YFP and YFP-H148Q) at the buffer pH, as well as the optical density at pH 9 in the absence of interacting anions, were used in the Henderson-Hasselbalch equation to estimate the chromophore pK_a for each condition examined. Microscopic binding constants for anion binding to the protein were extracted by curve fitting of the chromophore pK_a to the anion concentration, using an expression for a linked equilibrium involving two different ligands.

The following examples are provided by way of illustration, not by way of limitation.

EXAMPLES

As a step in understanding the properties of GFP, and to aid in the tailoring of GFPs with altered characteristics, we have determined the three dimensional structure at 1.9 Å resolution of the S65T mutant (R. Heim et al. *Nature* 373:664-665 (1995)) of *A. victoria* GFP. This mutant also contains the ubiquitous Q80R substitution, which accidentally occurred in the early distribution of the GFP cDNA and is not known to have any effect on the protein properties (M. Chalfie et al. *Science* 263:802-805 (1994)).

Histidine-tagged S65T GFP (R. Heim et al. *Nature* 373:664-665 (1995)) was overexpressed in JM109/pRSET$_B$ in 4 l YT broth plus ampicillin at 37°C, 450 rpm and 5 l/min air flow. The temperature was reduced to 25°C at A_{595} = 0.3, followed by induction with 1mM isopropylthiogalactoside for 5h. Cell paste was stored at –80°C

overnight, then was resuspended in 50 mM HEPES pH 7.9, 0.3 M NaCl, 5 mM 2-mercaptoethanol, 0.1 mM phenylmethyl-sulfonylfluoride (PMSF), passed once through a French press at 10,000 psi, then centrifuged at 20 K rpm for 45 min. The supernatant was applied to a Ni-NTA-agarose column (Qiagen), followed by a wash with 20 mM imidazole, then eluted with 100 mM imidazole. Green fractions were pooled and subjected to chymotryptic (Sigma) proteolysis (1:50 w/w) for 22 h at RT. After addition of 0.5 mM PMSF, the digest was reapplied to the Ni column. N-terminal sequencing verified the presence of the correct N-terminal methionine. After dialysis against 20 mM HEPES, pH 7.5 and concentration to $A_{490} = 20$, rod-shaped crystals were obtained at RT in hanging drops containing 5 µl protein and 5 µl well solution, 22-26% PEG 4000 (Serva), 50 mM HEPES pH 8.0-8.5, 50 mM $MgCl_2$ and 10 mM 2-mercapto-ethanol within 5 days. Crystals were 0.05 mm across and up to 1.0 mm long. The space group is $P2_12_12_1$ with a = 51.8, b = 62.8, c = 70.7 Å, Z=4. Two crystal forms of wild-type GFP, unrelated to the present form, have been described by M. A. Perrozo, K. B. Ward, R. B. Thompson, & W. W. Ward. *J. Biol. Chem.* 203, 7713-7716 (1988).

The structure of GFP was determined by multiple isomorphous replacement and anomalous scattering (Table E), solvent flattening, phase combination and crystallographic refinement. The most remarkable feature of the fold of GFP is an eleven stranded ß-barrel wrapped around a single central helix (Fig. 1A and 1B), where each strand consists of approximately 9-13 residues. The barrel forms a nearly perfect cylinder 42 Å long and 24 Å in diameter. The N-terminal half of the polypeptide comprises three anti-parallel strands, the central helix, and then 3 more anti-parallel strands, the latter of which (residues 118-123) is parallel to the N-terminal strand (residues 11-23). The polypeptide backbone then crosses the "bottom" of the molecule to form the second half of the barrel in a five-strand Greek Key motif. The top end of the cylinder is capped by three short, distorted helical segments, while one short, very distorted helical segment caps the bottom of the cylinder. The main-chain hydrogen bonding lacing the surface of the cylinder very likely accounts for the unusual stability of the protein towards denaturation and proteolysis. There are no large segments of the polypeptide that could be excised

while preserving the intactness of the shell around the chromophore. Thus it would seem difficult to re-engineer GFP to reduce its molecular weight (J. Dopf & T.M. Horiagon *Gene* 173:39-43 (1996)) by a large percentage.

The *p*-hydroxybenzylideneimidazolidinone chromophore (C. W. Cody et al. *Biochemistry* 32:1212-1218 (1993)) is completely protected from bulk solvent and centrally located in the molecule. The total and presumably rigid encapsulation is probably responsible for the small Stokes' shift (i.e. wavelength difference between excitation and emission maxima), high quantum yield of fluorescence, inability of O_2 to quench the excited state (B.D. Nageswara Rao et al. *Biophys. J.* 32:630-632 (1980)), and resistance of the chromophore to titration of the external pH (W. W. Ward. *Bioluminescence and Chemiluminescence* (M. A. DeLuca and W. D. McElroy, eds) Academic Press pp. 235-242 (1981); W. W. Ward & S. H. Bokman. *Biochemistry* 21:4535-4540 (1982); W. W. Ward et al. *Photochem. Photobiol.* 35:803-808 (1982)). It also allows one to rationalize why fluorophore formation should be a spontaneous intramolecular process (R. Heim et al. *Proc. Natl. Acad. Sci.* USA 91:12501-12504 (1994)), as it is difficult to imagine how an enzyme could gain access to the substrate. The plane of the chromophore is roughly perpendicular (60) to the symmetry axis of the surrounding barrel. One side of the chromophore faces a surprisingly large cavity, that occupies a volume of approximately 135 $Å^3$ (B. Lee & F. M. Richards. *J. Mol. Biol.* 55:379-400 (1971)). The atomic radii were those of Lee & Richards, calculated using the program MS with a probe radius of 1.4 Å. (M. L. Connolly, *Science* 221:709-713 (1983)). The cavity does not open out to bulk solvent. Four water molecules are located in the cavity, forming a chain of hydrogen bonds linking the buried side chains of Glu^{222} and Gln^{69}. Unless occupied, such a large cavity would be expected to de-stabilize the protein by several kcal/mol (S. J. Hubbard et al., *Protein Engineering* 7:613-626 (1994); A. E. Eriksson et al. *Science* 255:178-183 (1992)). Part of the volume of the cavity might be the consequence of the compaction resulting from cyclization and dehydration reactions. The cavity might also temporarily accommodate the oxidant, most likely O_2 (A. B. Cubitt et al. *Trends Biochem. Sci.* 20:448-455 (1995); R. Heim et al. *Proc. Natl. Acad. Sci.* USA 91:12501-12504 (1994); S. Inouye & F.I. Tsuji. *FEBS Lett.* 351:211-214 (1994)), that

dehydrogenates the α-β bond of Tyr^{66}. The chromophore, cavity, and side chains that contact the chromophore are shown in Figure 2A and a portion of the final electron density map in this vicinity in 2B.

The opposite side of the chromophore is packed against several aromatic and polar side chains. Of particular interest is the intricate network of polar interactions with the chromophore (Fig. 2C). His^{148}, Thr^{203} and Ser^{205} form hydrogen bonds with the phenolic hydroxyl; Arg^{96} and Gln^{94} interact with the carbonyl of the imidazolidinone ring and Glu^{222} forms a hydrogen bond with the side chain of Thr^{65}. Additional polar interactions, such as hydrogen bonds to Arg^{96} from the carbonyl of Thr^{62}, and the side-chain carbonyl of Gln^{183}, presumably stabilize the buried Arg^{96} in its protonated form. In turn, this buried charge suggests that a partial negative charge resides on the carbonyl oxygen of the imidazolidinone ring of the deprotonated fluorophore, as has previously been suggested (W. W. Ward. *Bioluminescence and Chemiluminescence* (M. A. DeLuca and W. D. McElroy, eds) Academic Press pp. 235-242 (1981); W. W. Ward & S. H. Bokman. *Biochemistry* 21:4535-4540 (1982); W. W. Ward et al. *Photochem. Photobiol.* 35:803-808 (1982)). Arg^{96} is likely to be essential for the formation of the fluorophore, and may help catalyze the initial ring closure. Finally, Tyr^{145} shows a typical stabilizing edge-face interaction with the benzyl ring. Trp^{57}, the only tryptophan in GFP, is located 13 Å to 15 Å from the chromophore and the long axes of the two ring systems are nearly parallel. This indicates that efficient energy transfer to the latter should occur, and explains why no separate tryptophan emission is observable (D.C. Prasher et al. *Gene* 111:229-233 (1992). The two cysteines in GFP, Cys^{48} and Cys^{70}, are 24 Å apart, too distant to form a disulfide bridge. Cys^{70} is buried, but Cys^{48} should be relatively accessible to sulfhydryl-specific reagents. Such a reagent, 5,5'-dithiobis(2-nitrobenzoic acid), is reported to label GFP and quench its fluorescence (S. Inouye & F.I. Tsuji *FEBS Lett.* 351:211-214 (1994)). This effect was attributed to the necessity for a free sulfhydryl, but could also reflect specific quenching by the 5-thio-2-nitrobenzoate moiety that would be attached to Cys^{48}.

Although the electron density map is for the most part consistent with the proposed structure of the chromophore (D.C. Prasher et al. *Gene* 111:229-233 (1992);

C. W. Cody et al. *Biochemistry* 32:1212-1218 (1993)) in the *cis* [Z-] configuration, with no evidence for any substantial fraction of the opposite isomer around the chromophore double bond, difference features are found at >4 σ in the final (F_o-F_c) electron density map that can be interpreted to represent either the intact, uncyclized polypeptide or a carbinolamine (inset to Fig. 2C). This suggests that a significant fraction, perhaps as much as 30% of the molecules in the crystal, have failed to undergo the final dehydration reaction. Confirmation of incomplete dehydration comes from electrospray mass spectrometry, which consistently shows that the average masses of both wild-type and S65T GFP (31,086±4 and 31,099.5±4 Da, respectively) are 6-7 Da higher than predicted (31,079 and 31,093 Da, respectively) for the fully matured proteins. Such a discrepancy could be explained by a 30-35% mole fraction of apoprotein or carbinolamine with 18 or 20 Da higher molecular weight The natural abundance of ^{13}C and ^{2}H and the finite resolution of the Hewlett-Packard 5989B electrospray mass spectrometer used to make these measurements do not permit the individual peaks to be resolved, but instead yields an average mass peak with a full width at half maximum of approximately 15 Da. The molecular weights shown include the His-tag, which has the sequence MRGSHHHHHH GMASMTGGQQM GRDLYDDDDK DPPAEF (SEQ ID NO:5). Mutants of GFP that increase the efficiency of fluorophore maturation might yield somewhat brighter preparations. In a model for the apoprotein, the Thr^{65}-Tyr^{66} peptide bond is approximately in the α-helical conformation, while the peptide of Tyr^{66}-Gly^{67} appears to be tipped almost perpendicular to the helix axis by its interaction with Arg^{96}. This further supports the speculation that Arg^{96} is important in generating the conformation required for cyclization, and possibly also for promoting the attack of Gly^{67} on the carbonyl carbon of Thr^{65} (A. B. Cubitt et al. *Trends Biochem. Sci.* 20:448-455 (1995)).

The results of previous random mutagenesis have implicated several amino acid side chains to have substantial effects on the spectra and the atomic model confirms that these residues are close to the chromophore. The mutations T203I and E222G have profound but opposite consequences on the absorption spectrum (T. Ehrig et al. *FEBS Letters* 367:163-166 (1995)). T203I (with wild-type Ser^{65}) lacks

the 475 nm absorbance peak usually attributed to the anionic chromophore and shows only the 395 nm peak thought to reflect the neutral chromophore (R. Heim et al. *Proc. Natl. Acad. Sci.* USA 91:12501-12504 (1994); T. Ehrig et al. *FEBS Letters* 367:163-166 (1995)). Indeed, Thr^{203} is hydrogen-bonded to the phenolic oxygen of the chromophore, so replacement by Ile should hinder ionization of the phenolic oxygen. Mutation of Glu^{222} to Gly (T. Ehrig et al. *FEBS Letters* 367:163-166 (1995)) has much the same spectroscopic effect as replacing Ser^{65} by Gly, Ala, Cys, Val, or Thr, namely to suppress the 395 nm peak in favor of a peak at 470-490 nm (R. Heim et al. *Nature* 373:664-665 (1995); S. Delagrave et al. *Bio/Technology* 13:151-154 (1995)). Indeed Glu^{222} and the remnant of Thr^{65} are hydrogen-bonded to each other in the present structure, probably with the uncharged carboxyl of Glu^{222} acting as donor to the side chain oxygen of Thr^{65}. Mutations E222G, S65G, S65A, and S65V would all suppress such H-bonding. To explain why only wild-type protein has both excitation peaks, Ser^{65}, unlike Thr^{65}, may adopt a conformation in which its hydroxyl donates a hydrogen bond to and stabilizes Glu^{222} as an anion, whose charge then inhibits ionization of the chromophore. The structure also explains why some mutations seem neutral. For example, Gln^{80} is a surface residue far removed from the chromophore, which explains why its accidental and ubiquitous mutation to Arg seems to have no obvious intramolecular spectroscopic effect (M. Chalfie et al. *Science* 263:802-805 (1994)).

The development of GFP mutants with red-shifted excitation and emission maxima is an interesting challenge in protein engineering (A. B. Cubitt et al. *Trends Biochem. Sci.* 20:448-455 (1995); R. Heim et al. *Nature* 373:664-665 (1995); S. Delagrave et al. *Bio/Technology* 13:151-154 (1995)). Such mutants would also be valuable for avoidance of cellular autofluorescence at short wavelengths, for simultaneous multicolor reporting of the activity of two or more cellular processes, and for exploitation of fluorescence resonance energy transfer as a signal of protein-protein interaction (R. Heim & R.Y. Tsien. *Current Biol.* 6:178-182 (1996)). Extensive attempts using random mutagenesis have shifted the emission maximum by at most 6 nm to longer wavelengths, to 514 nm (R. Heim & R.Y. Tsien. *Current Biol.* 6:178-182 (1996)); previously described "red-shifted" mutants merely suppressed the

395 nm excitation peak in favor of the 475 nm peak without any significant reddening of the 505 nm emission (S. Delagrave et al. *Bio/Technology* 13:151-154 (1995)). Because Thr^{203} is revealed to be adjacent to the phenolic end of the chromophore, we mutated it to polar aromatic residues such as His, Tyr, and Trp in the hope that the additional polarizability of their systems would lower the energy of the excited state of the adjacent chromophore. All three substitutions did indeed shift the emission peak to greater than 520 nm (Table F). A particularly attractive mutation was T203Y/S65G/V68L/S72A, with excitation and emission peaks at 513 and 527 nm respectively. These wavelengths are sufficiently different from previous GFP mutants to be readily distinguishable by appropriate filter sets on a fluorescence microscope. The extinction coefficient, 36,500 $M^{-1}cm^{-1}$, and quantum yield, 0.63, are almost as high as those of S65T (R. Heim et al. *Nature* 373:664-665 (1995)).

Comparison of *Aequorea* GFP with other protein pigments is instructive. Unfortunately, its closest characterized homolog, the GFP from the sea pansy *Renilla reniformis* (O. Shimomura and F.H. Johnson *J. Cell. Comp. Physiol.* 59:223 (1962); J. G. Morin and J. W. Hastings, *J. Cell. Physiol.* 77:313 (1971); H. Morise et al. *Biochemistry* 13:2656 (1974); W. W. Ward *Photochem. Photobiol. Reviews* (Smith, K. C. ed.) 4:1 (1979); W. W. Ward. *Bioluminescence and Chemiluminescence* (M. A. DeLuca and W. D. McElroy, eds) Academic Press pp. 235-242 (1981); W. W. Ward & S. H. Bokman *Biochemistry* 21:4535-4540 (1982); W. W. Ward et al. *Photochem. Photobiol.* 35:803-808 (1982)), has not been sequenced or cloned, though its chromophore is derived from the same FSYG sequence as in wild-type *Aequorea* GFP (R. M. San Pietro et al. *Photochem. Photobiol.* 57:63S (1993)). The closest analog for which a three dimensional structure is available is the photoactive yellow protein (PYP, G. E. O. Borgstahl et al. *Biochemistry* 34:6278-6287 (1995)), a 14-kDa photoreceptor from halophilic bacteria. PYP in its native dark state absorbs maximally at 446 nm and transduces light with a quantum yield of 0.64, rather closely matching wild-type GFP's long wavelength absorbance maximum near 475 nm and fluorescence quantum yield of 0.72-0.85. The fundamental chromophore in both proteins is an anionic *p*-hydroxycinnamyl group, which is covalently attached to the protein via a thioester linkage in PYP and a heterocyclic iminolactam in GFP. Both

proteins stabilize the negative charge on the chromophore with the help of buried cationic arginine and neutral glutamic acid groups, Arg^{52} and Glu^{46} in PYP and Arg^{96} and Glu^{222} in GFP, though in PYP the residues are close to the oxyphenyl ring whereas in GFP they are nearer the carbonyl end of the chromophore. However, PYP has an overall α/β fold with appropriate flexibility and signal transduction domains to enable it to mediate the cellular phototactic response, whereas GFP is a much more regular and rigid β-barrel to minimize parasitic dissipation of the excited state energy as thermal or conformational motions. GFP is an elegant example of how a visually appealing and extremely useful function, efficient fluorescence, can be spontaneously generated from a cohesive and economical protein structure.

A. Summary Of GFP Structure Determination

Data were collected at room temperature in house using either Molecular Structure Corp. R-axis II or San Diego Multiwire Systems (SDMS) detectors (Cu Kα) and later at beamline X4A at the Brookhaven National Laboratory at the selenium absorption edge (λ = 0.979 Å) using image plates. Data were evaluated using the HKL package (Z. Otwinowski, in *Proceedings of the CCP4 Study Weekend: Data Collection and Processing*, L. Sawyer, N. Issacs, S. Bailey, Eds. (Science and Engineering Research Council (SERC), Daresbury Laboratory, Warrington, UK, (1991)), pp 56-62; W. Minor, XDISPLAYF (Purdue University, West Lafayette, IN, 1993)) or the SDMS software (A. J. Howard et al. *Meth. Enzymol.* 114:452-471 (1985)). Each data set was collected from a single crystal. Heavy atom soaks were 2 mM in mother liquor for 2 days. Initial electron density maps were based on three heavy atom derivatives using in-house data, then later were replaced with the synchrotron data. The EMTS difference Patterson map was solved by inspection, then used to calculate difference Fourier maps of the other derivatives. Lack of closure refinement of the heavy atom parameters was performed using the Protein package (W. Steigemann, in Ph.D. Thesis (Technical University, Munich, 1974)). The MIR maps were much poorer than the overall figure of merit would suggest, and it was clear that the EMTS isomorphous differences dominated the phasing. The enhanced anomalous occupancy for the synchrotron data provided a partial solution to

the problem. Note that the phasing power was reduced for the synchrotron data, but the figure of merit was unchanged. All experimental electron density maps were improved by solvent flattening using the program DM of the CCP4 (CCP4: *A Suite of Programs for Protein Crystallography* (SERC Daresbury Laboratory, Warrington WA4 4AD UK, 1979)) package assuming a solvent content of 38%. Phase combination was performed with PHASCO2 of the Protein package using a weight of 1.0 on the atomic model. Heavy atom parameters were subsequently improved by refinement against combined phases. Model building proceeded with FRODO and O (T. A. Jones et al. *Acta. Crystallogr. Sect. A* 47:110 (1991); T. A. Jones, in *Computational Crystallography* D. Sayre, Ed. (Oxford University Press, Oxford, 1982) pp. 303-317) and crystallographic refinement was performed with the TNT package (D. E. Tronrud et al. *Acta Cryst. A* 43:489-503 (1987)). Bond lengths and angles for the chromophore were estimated using CHEM3D (Cambridge Scientific Computing). Final refinement and model building was performed against the X4A selenomethione data set, using ($2F_o$-F_c) electron density maps. The data beyond 1.9 Å resolution have not been used at this stage. The final model contains residues 2-229 as the terminal residues are not visible in the electron density map, and the side chains of several disordered surface residues have been omitted. Density is weak for residues 156-158 and coordinates for these residues are unreliable. This disordering is consistent with previous analyses showing that residues 1 and 233-238 are dispensable but that further truncations may prevent fluorescence (J. Dopf & T.M. Horiagon. *Gene* 173:39-43 (1996)). The atomic model has been deposited in the Protein Data Bank (access code 1EMA).

Table G

Diffraction Data Statistics

Crystal	Resolution (Å)	Total obs	Unique obs	Compl. (%)[a]	Compl. (shell)[b]	Rmerge (%)[c]	Riso (%)[d]
R-axix II							
Native	2.0	51907	13582	80	69	4.1	5.8
EMTS[e]	2.6	17727	6787	87	87	5.7	20.6
SeMet	2.3	44975	10292	92	88	10.2	9.3
Multiwire							
HGI4-Se	3.0	15380	4332	84	79	7.2	28.8
X4a							
SeMet	1.8	126078	19503	80	55	9.3	9.4
EMTS	2.3	57812	9204	82	66	7.2	26.3

Phasing Statistics

Derivative	Resolution (Å)	Number of sites	Phasing power[f]	Phasing Power(shell)	FOM[g]	FOM (shell)
In House						
EMTS	3.0	2	2.08	.2.08	0.77	.072
SeMet	3.0	4	1.66	1.28	-	-
HGI4-Se	3.0	9	1.77	1.90	-	-
X4a						
EMTS	3.0	2	1.36	1.26	0.77	.072
SeMet	3.0	4	1.31	1.08	-	-

Atomic Model Statistics

Protein atoms	1790
Solvent atoms	94
Resol. range (Å)	20-1.9
Number of reflections (F > 0)	17676
Completeness	84.
R. factor[(h)]	0.175
Mean B-value ($Å^2$)	24.1
Deviations from ideality	
Bond lengths (Å)	0.014
Bond angles (°)	1.9
Restrained B-values ($Å^2$)	4.3
Ramachandran outliers	0

Notes:

(a) Completeness is the ratio of observed reflections to theoretically possible expressed as a percentage.

(b) Shell indicates the highest resolution shell, typically 0.1-0.4 Å wide.

(c) Rmerge = $\Sigma |I - \langle I \rangle| / \Sigma I$, where $\langle I \rangle$ is the mean of individual observations of intensities I.

(d) Riso = $\Sigma |I_{DER} - I_{NAT}| / \Sigma I_{NAT}$

(e) Derivatives were EMTS=ethymercurithiosalicylate (residues modified Cys^{48} and Cys^{70}), SeMet=selenomethionine substituted protein (Met^{1} and Met^{233} could not be located); HgI_4-SeMet = double derivative HgI_4 on SeMet background.

(f) Phasing power = $\langle F_H \rangle / \langle E \rangle$ where $\langle F_H \rangle$=r.m.s. heavy atom scattering and $\langle E \rangle$=lack of closure.

(g) FOM, mean figure of merit

(h) Standard crystallographic R-factor, $R = \Sigma ||F_{obs}| - |F_{calc}|| / \Sigma |F_{obs}|$

B. Spectral properties of Thr^{203} ("T203") mutants compared to S65T

The mutations F64L, V68L and S72A improve the folding of GFP at 37° (B. P. Cormack et al. *Gene* 173:33 (1996)) but do not significantly shift the emission spectra.

TABLE H

Clone	Mutations	Excitation max.(nm)	Extinction coefficient ($10^3 M^{-1}cm^{-1}$)	Emission max.(nm)
S65T	S65T	489	39.2	511
5B	T203H/S65T	512	19.4	524
6C	T203Y/S65T	513	14.5	525

10B	T203Y/F64L/S65G/S72A	513	30.8	525
10C	T203Y/F65G/V68L/S72A	513	36.5	527
11	T203W/S65G/S72A	502	33.0	512
12H	T203Y/S65G/S72A	513	36.5	527
20A	T203Y/S65G/V68L/Q69K/S72A	515	46.0	527

C. YFP and YFP-H148Q as halide sensors at acidic and neutral pH

The absorbance spectrum of YFP is a function of NaCl concentration (Fig.7), with conversion of band B, the chromophore anion (λmax 514 nm), to band A, the neutral form (λmax 392 nm) upon addition of chloride. Since only the anion is fluorescent in the YFPs, suppression of fluorescence occurs concomitant with increasing [NaCl]. For YFP, a clean isosbestic point is observed (Fig. 7), whereas for YFP-H148Q, the isosbestic point is less well defined (Jayaraman *et al.*, 2000) . The effect is fully reversible. In YFP-H148Q and YFP-H148G, the absorbance maximum of band A is blue-shifted by 20 nm (from 415 nm to 395 nm) upon addition of salt, though the absorbance maximum of band B is unaffected. The detailed binding equilibria and anion specificities are discussed below.

To further establish the usefulness of the YFPs as halide sensors, in particular for organelles that are more acidic than the cytosol (pH 7.4) where CFTR pumping was assayed (Jayaraman *et al.*, 2000) , emission intensity of YFP and YFP-H148Q was measured between 0 and 150 mM NaCl at pH 6.0 to 8.0 (Fig. 8A,B), under conditions of constant ionic strength. We found that YFP constitutes an excellent probe under acidic conditions. At pH 6.0, fluorescence decreased by 39% from 0 to 20 mM NaCl, whereas at the cytosolic pH of 7.5, the decrease is only 3.2% under identical conditions. For YFP-H148Q titrated with NaCl, fluorescence loss is also large at pH 6.0 (48%), and remains fairly significant (11%) at pH 7.5. For measurements of chloride concentrations in the low millimolar range near neutral pH,

YFP-H148Q appears to be the preferred probe. If iodide is substituted for chloride, YFP-H148Q fluorescence loss is much larger, even at pH 7.5 where a loss of 50% is observed (0 to 20 mM NaI) (Fig. 8B). This observation has recently been exploited in vivo, in studies of Cl^-/I^- exchange by the CFTR channel in plasma membranes (Jayaraman *et al.*, 2000) . In contrast to the above variant, in the original YFP the magnitude of the iodide effect is more comparable to the chloride effect (see binding data below).

D. Crystallographic identification and description of halide binding sites

We determined two crystal structures of YFP-H148Q, one containing two bound iodides (100 mM iodide soak), and the other containing no bound halides at all (200 mM chloride soak). The respective R-factors are 18.8% and 20.4% to a resolution of 2.1 Å, and the geometry is reasonably good. A summary of the relevant crystallographic statistics is presented in Table I. Since iodine has an anomalous signal at the in-house CuKα wavelength, an anomalous difference map was calculated for the iodide soak in order to identify heavy atom positions. We found two distinct electron density peaks at 7.7 and 5.5 rms deviations respectively, one located close to the chromophore and buried in the protein interior, the other in a small indentation on the protein surface near Trp57 at the cap of the barrel (Data not shown).

The buried iodide refined to an occupancy of 0.60 with the thermal factor fixed at 30 $Å^2$, indicating that binding in the crystal at 100 mM iodide is not nearly as tight as in solution, where the binding constant is 2.7 mM (see below). This iodide is located 4.3 Å away from the chromophore heterocyclic carbonyl oxygen, and is involved in a charge interaction with Arg96, with a distance of 4.1 Å to NE2 of the guanidinium group (Fig. 9 and 10), a buried positive charge that is likely providing a large fraction of the anion binding energy. Furthermore, the iodide is hydrogen bonded to both the phenolic hydroxyl of Tyr203 and the side chain amide nitrogen of Gln69, with hydrogen bonding distances of 3.3 and 3.2 Å respectively (Fig. 9 and 10). These distances are within the range of hydrogen bonding distances expected for iodide interacting with oxygen or nitrogen. A statistical database analysis of small

molecule crystal structures found that the mean distance between iodide and ā phenolic hydroxyl is 3.47 Å, and between iodide and a sp^2-hybridized nitrogen is 3.66 Å (Steiner, 1998) . In a crystal structure of haloalkane dehalogenase with bound iodide, it was found that the iodide is 3.4 and 3.6 Å away from the indole nitrogens of two tryptophans, 3.3 Å from a solvent molecule, and 4.2 Å from the phenolic oxygen of a tyrosine (Verschueren *et al.*, 1993).

The buried halide also interacts with the aromatic rings of the chromophore and Tyr203 (Fig. 9). Anions are often preferentially located in or near the plane of aromatic rings, since aromatic ring hydrogens carry a partial positive charge (Burley & Petsko, 1988). In YFP-H148Q, the iodide is not quite in the plane of either of the two π systems, but is located roughly equidistant from the 2 planes, offset from the center of the stacking interaction, 4.1 Å from the aromatic CE1 of Tyr203 and 4.5 Å from the aromatic CD2 of the chromophore (Fig. 9). On the opposite side of the binding site, a series of hydrophobic residues line the halide binding site, consisting of Ile152, Leu201, Val163, Val150, and Phe165, all near van der Waals contact with the iodide. Both aromatic edge interactions with tyrosines and tryptophans, as well as apolar interactions with hydrophobic side chains are commonly found in other halide binding sites in proteins, such as in haloalkane dehalogenases (Pikkemaat *et al.*, 1999).

The second, surface-bound iodide is hydrogen-bonded to the amide nitrogen of Trp 57 and several ordered solvent molecules. This exposed anion is 16 Å from the chromophore phenolic oxygen, indicating that its influence on the chromophore charge state is negligible. The occupancy of this halide refines to 0.41 with the B-factor fixed at 39 $Å^2$, consistent weak binding as compared to the primary iodide adjacent to the chromophore.

E. Conformational changes adjacent to the buried iodide

The anion-binding pocket near the chromophore appears to be empty in the apo-structure of YFP-H148Q, in spite of the fact that the crystals were grown in the presence of 180 mM chloride, followed by soaking in 200 mM chloride. The solution

binding constant of 28 mM for chloride predicts that most of the pocket is occupied by chloride at the pH of the crystal mother liquor, 4.6 (see below). As has been found with the iodide soak, the molecules in the crystals do not appear to bind anions as tightly as in solution, possibly due to crystal packing forces. The volume of the internal cavity in YFP-H148Q is 55 $Å^3$, calculated using a probe with a radius of 1.2 Å (Connolly, 1985).

The iodide-containing cavity in the bound structure of YFP-H148Q is larger, 91 $Å^3$, to accommodate the rather large iodide which has a van der Waals volume of 42 $Å^3$. A series of conformational changes of various side chains lining the pocket are observed (Fig. 11). The largest movement is observed by Gln69, where the side chain amide has swung out from the center of the cavity, resulting in a 2.6 Å movement of the NE2 which is hydrogen-bonded to the halide (Fig. 11 and 12). Gln183 NE2 has moved out by 1.0 Å, though it is not clear whether it is a hydrogen bond donor to the iodide or Gln94 (NE2 and OE may be assigned oppositely). The apolar side chains of Leu201, Ile152, Val150, and Val163 (Fig. 11) all undergo movements to increase the cavity size in the presence of iodide, with their terminal carbons (CD1 for leucine and isoleucine, CG1 for valines) shifting by 2.4 Å, 1.9 Å, 1.6 Å, and 1.2 Å respectively. The aromatic ring plane of Phe165 has rotated by about 25°.

The phenolic hydroxyl of Tyr203 has shifted towards the iodide-containing cavity by 0.6 Å, likely to improve the hydrogen bonding interaction with the halide (Fig.12). There appears to be some flexibility in positioning the Tyr203 side chain next to the chromophore, presumably since it protrudes into a large water-filled cavity originally identified in the structure of GFP S65T (Ormo *et al.*, 1996) . An C_α-carbon overlay of 5 structures of YFP and its variants (Wachter *et al.*, 1998) shows that the C_βs of Tyr203 overlay quite well, whereas the phenolic oxygen varies by up to 1.4 Å. The hydrogen bond between Tyr203 and the halide appears to be of major importance in the generation of a halide binding site with reasonably tight affinity (see mutational analysis below). The chromophore shift toward the halide may also serve to improve aromatic edge interactions with the anion. As a consequence, the carboxylate of Glu222 has rotated away from the chromophore ring nitrogen (distance

increases from 3.3 to 3.6 Å), and is now involved in a tight hydrogen bond to Ser205 (Fig.12).

F. Relationship between anion binding and cavity size

The buried iodide site in YFP-H148Q identifies a small cavity that is present in a number of structures examined and does not vary much in size (Figure 13). Calculating van der Waals volumes using a sphere with a probe radius of 1.2 Å (Connolly, 1985) , the volume of this cavity is 21 $Å^3$ in WT GFP (Brejc *et al.*, 1997) , 19 $Å^3$ in GFP S65T (Ormo *et al.*, 1996) , 16 $Å^3$ in YFP and YFP-H148G (Wachter *et al.*, 1998) , and 21 $Å^3$ in YFP-H148G soaked in 500 mM KBr, where the crystallographic analysis shows that the binding site is also empty (unpublished data). The position of these cavities is essentially the same in the GFPs listed above, with its center close to Val150, Val163, Leu201, Ile152, Gln183, and Gln69, but about 6.6 Å distant from the chromophore methylene bridge and 6.1 Å from Arg96. WT (see below) and S65T GFP (Wachter & Remington, 1999) do not appear to interact with NaCl. On the other hand, all YFPs examined show anion interactions, with tightest Cl^- binding observed for YFP (see below). Clearly, cavity size and position are not directly correlated with Cl^- binding.

The cavities described above are too small to bind chloride, bromide, or iodide, whose van der Waals volumes range from 24.8 to 54 $Å^3$. Conformational changes are clearly necessary to allow for the interaction with any anions. In the apo-structure of YFP-H148Q, the cavity is somewhat larger, even in the absence of bound anions, with a volume of 55 $Å^3$. In this variant, the cavity is extended towards the chromophore, and the volume is increased by small movements of side chain atoms (0.4 Å and 1.2 Å) lining the binding site (Gln69, Tyr203, Val150, Val163, Phe165, Arg96, His181), and the chromophore itself. Many of these residues undergo further shifts upon iodide binding, as described above and figure 6. Compensating movements of the terminal side-chain carbons of Ile152 (2.1 Å) and Leu201 (2.3 Å) lead to some repacking of the hydrophobic core, without changing the adjacent cavity

surface much. The larger cavity size of YFP-H148Q may in part be responsible for the unexpectedly tight binding of iodide compared to chloride.

G. Relaxation of the β-barrel in response to the H148Q substitution and iodide binding

Both the introduction of H148Q in YFP background in the absence of halides, and the binding of iodide to YFP-H148Q lead to structural adjustments of β-strands 7 (residues 143 to 154) and 8 (residues 160 to 171). These adjustments are evident from C_α overlays of YFP and YFP-H148Q, with an rms deviation of 0.42 Å, and of YFP-H148Q with and without I^-, with an rms deviation of 0.47 Å (Figure 14). At one cap of the barrel, strands 7 and 8 are connected via a turn centered on residue 158, whereas near the other cap of the barrel, strand 7 forms a β-bulge around residue 148, and main chain β-sheet interactions are disrupted (Ormo et al., 1996; Wachter et al., 1998). Instead, several ordered solvent molecules and side-chain contacts (His148 in YFP, Gln148 in YFP-H148Q) form a hydrogen bond network between the strands (Figure 14). Upon substitution of His148 with Gln, the α-carbon of residue 166 is pulled in towards the center of the barrel by 0.94 Å, and the α-carbon of residue 148 is pushed out by 0.94 Å. These movements are compensated for by structural adjustments within the adjacent loop regions (1.4 Å shift by the α-carbon of residue 172, and 0.94 Å by the α-carbon of residue 157). None of these residues are involved in crystal contacts in either of the two structures.

Upon binding of iodide to YFP-H148Q, Lys166 is pushed back out from the center of the barrel by 1.0 Å, and is located near its original position in YFP (Figure 14). Likely, this shift in position occurs in response to the expansion of the buried cavity. Lys166 is not involved in a crystal contact in either of the two structures, whereas nearby Arg168 forms an intermolecular salt bridge with Asp149 upon iodide binding, but not without. The hydrogen bond of the Lys166 backbone oxygen to the side chain of Gln148 is not disrupted by this movement. Compensating shifts are

again observed at the end of this strand, where the backbone loop residues 172 and 173 are pulled in by up to 1.7 Å, though density in this area is less well defined. Whether halide binding to YFP has a similar effect on the β–bulge region as in YFP-H148Q is not known. Backbone movements in position 148 have been observed previously in YFP-H148G (Wachter *et al.*, 1998), consistent with increased flexibility in that part of the barrel.

H. Solvent accessibility of the chromophore in YFP-H148Q

The structure of YFP-H148Q shows that the Gln148 side chain is swung out towards the protein exterior (Figure 9), unlike the original histidine imidazole that is hydrogen bonded to the chromophore hydroxyl (Wachter *et al.*, 1998) , and constitutes a barrier to bulk solvent. Even before a structure was available, we predicted that Gln148 may be flipped out into the solvent (Elsliger *et al.*, 1999) , since partial chromophore exposure to exterior solvent may explain the higher pK_a of YFP-H148Q as compared to YFP (see Table K). Both in the apo and iodide-bound structure of YFP-H148Q, the Gln148 side chain amide nitrogen NE2 is hydrogen-bonded to the backbone carbonyl oxygen of Lys166, and the amide oxygen OE1 to the backbone nitrogen of Asn149 (Fig. 12 and 14), well away from the chromophore. Calculations of solvent-accessible surface (Connolly, 1983) using a probe sphere radius of 1.4 Å, as implemented by MidasPlus™, show that a shallow invagination on the protein surface is formed where the wild-type imidazole of His148 was located (Fig. 13). This solvent pocket is nearly in contact with the chromophore van der Waals surface. If one considers protein breathing motions, some solvent access that is not observable in the crystal structure is likely to occur. As compared to YFP-H148G (Wachter *et al.*, 1998) , where the solvent channel is directly in contact with the chromophore cavity, the channel of YFP-H148Q is truncated, consistent with triplet state photobleaching experiments which suggested that the chromophore is not exposed to aqueous-phase quenchers (Jayaraman *et al.*, 2000) .

I. Energetic analysis of linkage between anion and proton binding

The strong dependence of chromophore pKa on specific anion binding can be described by a linked binding equilibrium that considers the interaction between two different ligands, the anion that binds adjacent to Arg96, and the proton that binds to the phenolic end of the chromophore. Positive cooperativity is indicated by the fact that binding of the anion facilitates binding of the proton, raising the pK_a of the chromophore. The binding constant for anion binding is therefore influenced by the amount of proton binding, and vice versa. Hence, in a simple system with one binding site each for two different ligands, one can define two microscopic binding constants, k_1 for anion binding when the proton is on and k_2 for anion binding when the proton is off. Our crystallographic analysis for YFP-H148Q is consistent with one relevant binding site for the anion, and a previous crystallographic analysis on S65T is consistent with one proton binding site on the chromophore (Elsliger *et al.*, 1999) . The observed extent of anion binding is a function of pH, hence the macroscopic binding constants lie somewhere between the limiting values of k_1 and k_2.

A mathematical description has been developed by J. Wyman (1964) and is presented by Cantor and Schimmel in Biophysical Chemistry, Part III, (Cantor & Schimmel, 1980) . Here, we apply the general equation 15-79 to the special case of having one binding site for each ligand, with pK_a° representing the chromophore pK_a in the absence of any bound anions:

$$pK_a = \log \{(k_1 + [\text{chloride}]) / k_1\} - \log \{(k_2 + [\text{chloride}]) / k_2\} + pK_a^\circ$$

equation (1)

Using absorbance measurements at both pH 6.5 and pH 7.0, the pK_a of YFP was determined as a function of anion concentration for a large number of different ions. The concentration of the particular anion of interest was varied between 0 and at least 150 mM (for the halides as high as 400 mM, Fig. 15), and ionic strength was controlled by the addition of potassium gluconate which does not interact with the YFPs (Wachter & Remington, 1999) . Results for interacting anions were fit to

equation (1), and the microscopic binding constants k_1, and k_2 where possible, were extracted from the curve fit (Figure 15 and Table J). In general, small monovalent anions appear to show some interaction with YFP. Binding is tightest for fluoride, with $k_1 = 0.214$ mM. Other anions that were found to interact, including the other halides, have microscopic binding constants in the low millimolar range, with trifluoroacetic acid (TFA) giving the weakest interaction in this series ($k_1 = 21.2$ mM).

There does not appear to be any particular molecular shape dependence for this interaction, since triatomic linear (e.g. thiocyanate), square planar molecules (e.g. perchlorate), trigonal (e.g. nitrite), and spherical (e.g. halides) molecules are also found to bind. Formate modulated the chromophore pK_a as well ($k_1 = 7.47$ mM), though a previous, somewhat preliminary experiment by fluorescence indicated no interaction (Wachter & Remington, 1999) . As expected, anion binding to the anionic chromophore is unfavorable, with k_2 in the high millimolar or in the molar range, often outside the range of measurement (Table J).

Table J. Anion binding to the YFP chromophore in order of decreasing interaction strength.[a]

Interacting anions	k_1 (mM)[b]	k_2 (mM)[b]
Fluoride F^-	0.214 (0.009)	301 (64)
Thiocyanate SCN^-	1.37 (0.02)	large[c]
Perchlorate ClO_4^-	1.46 (0.36)	175 (11)
Nitrite NO_2^-	2.12 (0.40)	273 (200)
Iodide I^-	2.46 (0.11)	325 (64)
Nitrate NO_3^-	4.44 (0.25)	large[c]
Chloride Cl^-	4.69 (0.17)	288 (40)

Formate $HCOO^-$	7.47 (0.004)	large[c]
Bromide Br^-	7.76 (1.00)	280 (126)
TFA CF_3COO^-	21.2 (3.7)	large[c]

[a]Conjugate bases (prevalent ion at pH 6 to 7) are listed in order of decreasing interaction strength.

[b]The numbers in parenthesis are a lower estimate of the standard deviation as determined by Kaleidagraph™.

[c]These binding constants could not be determined since they fall outside the range of measurement, and are likely in the molar range.

Divalent anions such as phosphate and sulfate, and larger monovalent anions such as gluconate, Good buffers (e.g. HEPES, PIPES), isethionate (2-hydroxyethanesulfonic acid), and TCA (trichloroacetic acid), do not interact (Table J), as indicated by a constant pK_a of about 5.4 for YFP, essentially the same as when measurements are carried out in low ionic strength buffers without addition of salts (Wachter & Remington, 1999) . Somewhat smaller monovalent anions that do not interact include phosphoric acid, bicarbonate, and acetate. The hydration energy may be of importance in discrimination of anions, since acetate is strongly solvated, whereas TCA is only weakly hydrated in aqueous solvents (March, 1992) . The series presented for YFP in Table J is very similar to the one determined for YFP-H148Q by fluorescence at pH 7.5 (Jayaraman *et al.*, 2000) , with only minor differences in ordering. For example, YFP-H148Q binds Br^- more strongly than Cl, whereas for YFP, the order is reversed, likely due to the larger binding site in YFP-H148Q.

J. Identification of key residues for anion binding by mutational analysis

To identify which substitutions in YFP (S65G/V68L/S72A/T203Y) are contributors to specific anion binding near the chromophore, we carried out a mutational analysis, converting the four substitutions back to wild-type one-by-one. We then determined the pK_a of these revertants in the absence of interacting anions, and measured their affinity to chloride and iodide by pK_a determination as a function of halide concentration, followed by curve fitting to equation (1). Revertant 1 (S65G/S72A/T203Y) and revertant 2 (S65G/V68L/S72A) exhibited well-behaved pH and halide titration behavior as is observed for the YFPs, and their pK_a and k_1 for chloride and iodide binding are compared with those obtained for the YFPs in Table K. Reversion of residue 68 or residue 203 raises the chromophore pKa to 5.8 and 6.4 respectively. Reversion of residue 68 leads to a slight loss of chloride affinity (k_1 = 13.2 mM, as compared to 4.69 mM in YFP), whereas reversion of residue 203 dramatically weakens the interaction (k_1 = 153 mM). As is evident from Table K, chloride affinity is strongly coupled to chromophore pKa, with a weakening of the anion interaction with increasing pKa.

Table K Microscopic dissociation constants for chloride and iodide binding to the YFPs and its revertants.

variant	substitutions	k_1 (mM)[a] for Cl^-	k_1 (mM)[a] for I^-	pK_a[b]
YFP	S65G/V68L/S72A/T203Y	4.69 (0.17)	2.46 (0.11)	5.4
revertant 1	S65G/S72A/T203Y	13.2 (0.34)	3.04 (0.11)	5.8
YFP-H148Q	S65G/V68L/S72A/H148Q/T203Y	28.4 (5.1)	2.68 (0.11)	6.7
YFP-H148G	S65G/V68L/S72A/H148G/T203Y	82.8 (18.3)	15.73 (2.6)	7.5
revertant 2	S65G/V68L/S72A	153 (26)	117 (16)	6.4

[a]A lower estimate of the standard deviation (as reported by Kaleidagraph™) is given in parenthesis.

[b]The chromophore pK_a determined by absorbance in the absence of any interfering anions, such as chloride (buffered with either HEPES or PIPES, 150 mM gluconate).

Since the only exception to this rule is revertant 2, it appears that the correlation is intact only in the presence of T203Y. This substitution appears to be indispensable for strong anion interactions.

Iodide binding appears to be considerably tighter than chloride binding for all variants tested (Table K). Any correlation with chromophore pKa is weak at best. The relative selectivity of iodide over chloride is strongest for YFP-H148Q, followed by YFP-H148G. This may reflect the fact that iodide is a larger, softer ion than chloride, more difficult to fit into a small cavity unless the particular variant allows for structural relaxation of the β-barrel (see above).

Revertants 3 (S72A/T203Y) and 4 (T203Y) were more difficult to analyze, since their titration behavior is similar to WT GFP. Their absorbance spectra exhibit a mixed ground state of bands A and B, and are nearly pH-independent above pH 6.5. Excitation of either band A or B leads to green fluorescence in these revertants, reminiscent of the excited-state deprotonation described for WT GFP (Chattoraj *et al.*, 1996). Addition of NaCl to 250 mM to revertants 3 and 4 at pH 6.5 changes the ratio of the two absorbance bands only to a small degree, resulting in roughly a 20% decrease of band B in favor of band A. In WT GFP at pH 6.5, no spectral change is observed upon addition of 250 mM NaCl under conditions of constant ionic strength, consistent with a sensitivity towards ionic strength (Ward *et al.*, 1982) but not specific anion binding.

The present invention provides novel long wavelength engineered fluorescent proteins. While specific examples have been provided, the above description is illustrative and not restrictive. Many variations of the invention will become apparent to those skilled in the art upon review of this specification. The scope of the invention should, therefore, be determined not with reference to the above description, but instead should be determined with reference to the appended claims along with their full scope of equivalents.

Publications

Baird, G. S., Zacharias, D. A. & Tsien, R. Y. (1999). Circular permutation and receptor insertion within green fluorescent proteins. *Proc. Natl. Acad. Sci. USA* **96**, 11241-11246.

Bell, A. F., He, X., Wachter, R. M. & Tonge, P. J. (2000). Probing the ground state structure of the green fluorescent protein chromophore using Raman spectroscopy. *Biochemistry* (in press).

Bokman, S. H. & Ward, W. W. (1981). Renaturation of Aequorea green fluorescent protein. *Biochem. Biophys. Res. Commun.* **101**, 1372-1380.

Born, M. (1920). Volumen und hydratationswaerme der ionen. *Z. Physik* **1**, 45-48.

Brejc, K., Sixma, T. K., Kitts, P. A., Kain, S. R., Tsien, R. Y., Ormö, M. & Remington, S. J. (1997). Structural basis for dual excitation and photoisomerization of the *Aequorea victoria* Green Fluorescent Protein. *Proc. Natl. Acad. Sci. USA.* **94**, 2306-2311.

Burley, S. K. & Petsko, G. A. (1988). Weakly polar interactions in proteins. *Adv. Protein Chem.* **39**, 125-189.

Cantor, C. R. & Schimmel, P. R. (1980). *Biophysical Chemistry Part III: The behavior of biological macromolecules*, W. H. Freeman and Company, New York, pp 866-878.

Chattoraj, M., King, B. A., Bublitz, G. U. & Boxer, S. G. (1996). Ultra-fast excited state dynamics in Green Fluorescent Protein: Multiple states and proton transfer. *Proc. Natl. Acad. Sci. USA* **93**, 8362-8367.

Collaborative Computational Project N. 4 (1994). The CCP4 Suite: Programs for protein crystallography. *Acta Cryst.* **D50**, 760-763.

Connolly, M. L. (1983). Solvent-accessible surfaces of proteins and nucleic acids. *Science* **221**(4612), 709-713.

Connolly, M. L. (1985). Computation of molecular volume. *J. Am. Chem. Soc.* **107**, 1118-1124.

De Giorgi, F., Ahmed, Z., Bastianutto, C., Brini, M., Jouvaille, L. S., Marsault, L. S., Murgia, M., Pinton, P., Pozzan, T. & Rizzutto, R. (1999). *Meth. Cell Biol.* **58**, 75-85.

Doi, N. & Yanagawa, H. (1999). Design of generic biosensors based on green fluorescent proteins with allosteric sites by directed evolution. *FEBS Lett.* **453**, 305-307.

Elsliger, M.-A., Wachter, R. M., Hanson, G. T., Kallio, K. & Remington, S. J. (1999). Structural and spectral response of green fluorescent protein variants to changes in pH. *Biochemistry* **38**, 5296-5301.

Feller, G., le Bussy, O., Houssier, C. & Gerday, C. (1996). Structural and functional aspects of chloride binging to Alteromonas haloplanctis alpha-amylase. *J. Biol. Chem.* **271**, 23836-23841.

Fersht, A. (1985). *Enzyme structure and mechanism*, W. H. Freeman and Company, New York, NY, pg 298.

Halm, D. R. & Frizzell, R. A. (1992). Anion permeation in an apical membrane chloride channel of a secretory epithelial cell. *J. Gen. Physiol.* **99**, 339-366.

Heim, R. (1999). Green fluorescent protein forms for energy transfer. *Methods Enzymol.* **302**, 408-423.

Heim, R., Prasher, D. C. & Tsien, R. Y. (1994). Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc. Natl. Acad. Sci. USA* **91**, 12501-12504.

Jayaraman, S., Haggie, P., Wachter, R. M., Remington, S. J. & Verkman, A. S. (2000). Mechanism and cellular applications of a green fluorescent protein-based halide sensor. *J. Biol. Chem.* (in press).

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjelgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. Sect. A* **47**, 110-119.

Kneen, M., Farinas, J., Li, Y. & Verkman, A. S. (1998). Green fluorescent protein as a noninvasive intracellular pH indicator. *Biophys. J.* **74**, 1591-1599.

Kraulis, P. (1991). MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.

Llopis, J., McCaffery, J. M., Miyawaki, A., Farquhar, M. & Tsien, R. Y. (1998). Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins. *Proc. Natl. Acad. Sci. USA* **95**, 6803-6808.

March, J. (1992). *Advanced Organic Chemistry*, John Wiley & Sons, New York, pg 272.

Matz, M. V., Fradkov, A. F., Labas, Y. A., Savitsky, A. P., Zaraisky, A. G., Markelov, M. L. & Lukyanov, S. A. (1999). Fluorescent proteins from nonbioluminescent Anthozoa species. *Nature Biotechnol.* **17**, 969-973.

Miesenbock, G., De Angelis, D. A. & Rothman, J. E. (1998). Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins. *Nature* **394**, 192-195.

Miyawaki, A., Griesbeck, O., Heim, R. & Tsien, R. Y. (1999). Dynamic and quantitative Ca2+ measurements using improved cameleons. *Proc. Natl. Acad. Sci. USA* **96**, 2135-2140.

Ormö, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y. & Remington, S. J. (1996). Crystal structure of the *Aequorea victoria* Green Fluorescent Protein. *Science* **273**, 1392-1395.

Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.

Palm, G. J. & Wlodawer, A. (1999). Spectral variants of green fluorescent protein. *Methods Enzymol.* **302**, 378-394.

Palm, G. J., Zdanov, A., Gaitanaris, G. A., Stauber, R., Pavlakis, G. N. & Wlodawer, A. (1997). The structural basis for spectral variations in green fluorescent protein. *Nat. Struct. Biol.* **4**(5), 361-365.

Pikkemaat, M. G., Ridder, I. S., Rozeboom, H. J., Kalk, K. H., Dijkstra, B. W. & Janssen, D. B. (1999). Crystallographic and kinetic evidence of a collision complex formed during halide import in haloalkane dehalogenase. *Biochemistry* **38**, 12052-12061.

Robey, R. B., Ruiz, O., Santos, A. V. P., Ma, J., Kear, F., Wang, L.-J., Li, C.-J., Bernardo, A. A. & Arruda, J. A. L. (1998). pH-Dependent fluorescence of a heterologously expressed Aequorea Green Fluorescent Protein mutant: In situ spectral characteristics and applicability to intracellular pH estimation. *Biochemistry* **37**, 9894-9901.

Scharnagl, C., Raupp-Kossmann, R. & Fischer, S. F. (1999). Molecular basis for pH sensitivity and proton transfer in Green Fluorescent Protein: Protonation and Conformational substates from electrostatic calculations. *Biophys. J.* **77**, 1839-1857.

Steiner, T. (1998). Hydrogen-bond distances to halide ions in organic and organometallic crystal structures: Up-to-date database study. *Acta Cryst.* **B54**, 456-463.

Sullivan, K. F. & Kay, S. A. E., eds (1999). Green fluorescent proteins. *Methods Cell Biol.* **58**, Academic Press, San Diego.

Topell, S., Hennecke, J. & Glockshuber, R. (1999). Circularly permuted variants of the green fluorescent protein. *FEBS Lett.* **457**, 283-289.

Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987). An efficient general-purpose least-squares refinement program for macromolecular structures. *Acta Crystallogr. Sect. A* **43**, 489-501.

Tsien, R. Y. (1998). The green fluorescent protein. *Ann. Rev. Biochem.* **67**, 509-544.

Verschueren, K. H. G., Kingma, J., Rozeboom, H. J., Kalk, K. H., Janssen, D. B. & Dijkstra, B. W. (1993). Crystallographic and fluorescence studies of the interaction of haloalkane dehalogenase with halide ions. Studies with halide compounds reveal a halide binding site in the active site. *Biochemistry* **32**, 9031-9037.

Wachter, R. M., Elsliger, M.-A., Kallio, K., Hanson, G. T. & Remington, S. J. (1998). Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein. *Structure* **6**(10), 1267-1277.

Wachter, R. M. & Remington, S. J. (1999). Sensitivity of the yellow variant of green fluorescent protein to halides and nitrate. *Curr. Biol.* **9**(17), R628-R629.

Wang, Z., Asenjo, A. B. & Oprian, D. D. (1993). Identification of the Cl^- binding site in the human red and green color vision pigments. *Biochemistry* **32**, 2125-2130.

Ward, W. W., Prentice, H. J., Roth, A. F., Cody, C. W. & Reeves, S. C. (1982). Spectral perturbations of the *Aequorea* green-fluorescent protein. *Photochem. Photobiol.* **35**, 803-808.

Wright, E. M. & Diamond, J. M. (1977). Anion selectivity in biological systems. *Physiological Reviews* **57**, 109-156.

Yang, F., Moss, L. G. & Phillips, G. N. (1996). The molecular structure of green fluorescent protein. *Nature Biotech.* **14**, 1246-1251.

Yuan, C., Kuwata, O., Liang, J., Misra, S., Balashov, S. P. & Ebrey, T. G. (1999). Chloride binding regulates the Schiff base pK in Gecko P521 cone-type visual pigment. *Biochemistry* **38**, 4649-4654.

All publications and patent documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication or patent document were so individually denoted.